STATISTICAL INFERENCE FOR SOME RISK MEASURES

A Dissertation
Presented to
The Academic Faculty

By

Yanxi Hou

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology

August 2017

STATISTICAL INFERENCE FOR SOME RISK MEASURES

Approved by:

Professor Xingxing Yu
Committee Chair
School of Mathematics
Georgia Institute of Technology

Professor Xingxing Yu, Co-Adivsor
School of Mathematics
Georgia Institute of Technology

Professor Liang Peng, Co-Adivsor
Department of Risk Management and Insurance
Robinson College of Business
Georgia State University

Professor Mayya Zhilova
School of Mathematics
Georgia Institute of Technology

Professor Galyna Livshyts
School of Mathematics
Georgia Institute of Technology

Professor Ionel Popescu
School of Mathematics
Georgia Institute of Technology

Date Approved: April 12, 2017

To my parents.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisors, Professors Liang Peng and Xingxing Yu, for their continuous support during my Ph.D. study, and for their patience, motivation, and immense knowledge. Professor Peng recruited me from Fudan University to be his Ph.D. student at Georgia Tech, and intrigued my interests in the study of empirical likelihood methods. Even after moving to Georgia State University, he still kept advising my research with patience and kindness. Professor Yu took great care of me and we are studying a new topic together to combine knowledge of our fields. I could not have imagined having better advisors and mentors for my Ph.D. study. In addition, I would also like to thank Professor Christian Houdré for introducting me into the several areas of probability, which widens and strengthens my research knowledge.

I would also like to thank Professors Ionel Popescu, Mayya Zhilova, and Galyna Livshyts, who serve on my thesis committee. In particular, I thank Professor Ionel Popescu and Professor Heinrich Matzinger, who served on my proposal committee, for their insightful comments and helpful questions on my research from various perspectives. I took two courses from Professor Ionel Popescu, which were very helpful for my research. I would also like to thank Professor Christopher Heil, who taught me real analysis courses, and Professor Vladimir Koltchinskii, who instructed the mathematical statistics courses. Their knowledge on these topics is impressive, intriguing my interests on the difficult part of analysis and encouraging me to work hard during my study.

I thank my coauthors for the stimulating discussions, for many sleepless nights we were working together on projects, and for the great ideas they shared with me: Yi He (Monash University) and Aiai Liu (Tongji University), for working jointly on empirical likelihood methods and topics in risk management; and Professor Alexandru V. Asimit and Russell Gerrard (City University London), for having a great discussion on a new tail dependence measure and solved a hard proof.

During the past four years, the life in Atlanta has been both enjoyable and fruitful. I would like to thank all my fellow students. Mr. Haiyu Zou is a very nice and humorous person who gave me tremendous help when I first arrived in Atlanta. Ms. Chengyin Jin was my classmate in the Statistics Department at Fudan University. It is really nice to have an old friend in Atlanta. She provided me with tips for living in Atlanta, including good supermarkets, great restaurants and other fun places. She and Haiyu taught me patiently how to drive. I would also like to thank Xing Wang and Xiaowei Yue, a lovely couple, who gave me a lot of help during the daily life.

Last but not the least, I would like to thank my parents, for supporting me spiritually throughout the writing of this thesis and my life in general. I am also grateful to my other family members and friends, who have supported me along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Recently, many risk measures have been developed for various types of risk based on multiple financial variables. However, statistical properties of these risk measures are not fully understood, and there are very few effective inference methods for them in applications to financial data. This thesis addresses asymptotic behaviors and statistical inference methods for several newly proposed risk measures, including relative risk and conditional value-at-risk. These risk metrics are intended to measure the tail risks and/or systemic risk in financial markets.

We consider conditional Value-at-Risk based on a linear regression model. We extend the assumptions on predictors and errors of the model, which make the model more flexible for the financial data. We then consider a relative risk measure based on a benchmark variable. The relative risk measure is proposed as a monitoring index for systemic risk of financial system. We also propose a new tail dependence measure based on the limit of conditional Kendall's tau. The new tail dependence can be used to distinguish between the asymptotic independence and dependence in extreme value theory.

For asymptotic results of these measures, we derive both normal and Chi-squared approximations. These approximations are a basis for inference methods. For normal approximation, the asymptotic variances are too complicated to estimate due to the complex forms of risk measures. Quantifying uncertainty is a practical and important issue in risk management. We propose several empirical likelihood methods to construct interval estimation based on Chi-squared approximation. Simulation study and real data analysis illustrate the usefulness of these risk measures and our inference methods. In particular, the empirical likelihood methods are very effective and easy to implement for practical applications.

CHAPTER 1
INTRODUCTION

My philosophy of statistics lies in the connection between two worlds, a true world that cannot be touched but can be inferred and an observed world that stays between the true world and our knowledge, expressing features of the true world to humans. The true world contains all the unknown information of randomness, including true distributions or parameters, which can never be surely known by human, unless we create it, like simulation. The observed world can be touched by human, and is the only path for man to explore and infer the true world. No sure thing exists since randomness covers the true world well, but useful knowledge and mathematical relationships can be derived by proper methods. Statistics is then the knowledge, accumulated by humans, of the inference for the connection between these two worlds. The ways to show the appropriateness of inference methods between the true and observed worlds are the most exciting and interesting part in statistics as I first studied it.

My dissertation addresses asymptotic results and statistical inference methods for some newly proposed risk measures. These risk measures were introduced to quantify the tail risks or systemic risk in financial markets. However, statistical properties of these risk measures are not yet fully understood and there are few effective inference methods in application to financial data. Here, we focus on both asymptotic results on the theoretical side and inference methods on the practical side.

In Chapter 2, we consider the conditional Value-at-Risk based on a linear regression model. We make some extensions on the assumptions on predictors and errors of the model which make the model more flexible for the financial data. Theorem 2.3.1 shows the asymptotic results of the conditional Value-at-Risk based on the profile empirical likelihood method. In Chapter 3, we consider a relative risk measure based on a benchmark variable. The relative risk measure is proposed as a monitoring index for systemic risk of a financial system. Theorem 3.2.3 and 3.3.1 show the asymptotic results for both i.i.d case and AR-GARCH case respectively. In Chapter 4, we propose a new tail dependence measure based on the limit of the conditional Kendall's tau. The new tail dependence can be used to distinguish between asymptotic independence and asymptotic dependence in extreme value theory which is stated in Theorem 4.2.2. In Chapter 5, we develop an interval estimation approach for the new tail dependence risk in Chapter 4 by Theorem 5.2.1.

In Section 1.1, we introduce the empirical likelihood method and its extensions which are the main inference approach in the thesis. In Section 1.2, we introduce the dependence concepts, including copulas and tail dependence. In Section 1.3, some well-known risk measures are mentioned which provides the basic components and ideas of complicated risk measures in the research.

1.1 Empirical Likelihood Methods

1.1.1 Likelihood-Based Methods

Likelihood methods are very classic and popular, which are widely applied in modern statistical inference approaches. They can be used to find efficient point estimators as well as to construct tests or confidence intervals with high coverage accuracy and good properties. They are very effective and flexible even when the data is distorted or incompletely observed or when there is incompatibility between the data and the model.

Parametric likelihood methods include assumptions on the joint distributions of the observed data: the joint distributions are of known forms with unknown parameters which we need to infer in the parametric space. When an assumption on the joint distributions is successful, parametric likelihood methods provide good statistical properties generally for both estimation and hypothesis tests. This advantage is an important factor that the likelihood-based methods can be applied widely in various fields without too many restrictions. However, this could also cause a problem: if the data sampled is far away from the assumed family of joint distributions, such likelihood-based methods could be inefficient; as a consequence, the approaches to construct confidence intervals or tests may completely fail.

Therefore, many statisticians turn to nonparametric inference methods which do not depend on the strong distributional assumptions. Although nonparametric approaches may not be as efficient as parametric ones, they give good asymptotic results. In practice, they are easy to handle, giving confidence intervals or tests with good properties similar to the parametric ones.

To illustrate the nonparametric maximum likelihood method, we assume a random variable $X \in \mathbb{R}$ with cumulative distribution function (CDF) F_0 such that $F_0(x) = \mathbb{P}(X \leq x)$ and $F_0(x^-) = \mathbb{P}(X < x)$, where $-\infty < x < \infty$. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed as X with the common distribution function F_0. We can estimate F_0 by the empirical cumulative distribution function (ECDF)

$$F_n(X) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x).$$

where $I(\cdot)$ is the indicator function. This gives a good point estimator and based on its normal asymptotic results, one can build tests or confidence interval. In addition, we are also interested in the nonparametric likelihood of a given CDF F

$$L(F) = \prod_{i=1}^{n}(F(X_i) - F(X_i^-)).$$

One idea to estimate F_0, the true CDF, is to maximize $L(F)$. This is because when the CDF is close to F_0, the observed data X_i may have the largest probability to be realized, which is similar to the idea in parametric likelihood methods. One interesting observation is that, for any CDF F,

$$L(F) < L(F_n), \quad \text{if} \quad F \neq F_n.$$

This means the ECDF is the nonparametric maximum likelihood estimator (NPMLE) of F_0. Thus, using F_n as a comparison, we can maximize instead the following nonparametric likelihood ratio function

$$\mathcal{R}(F) = \frac{L(F)}{L(F_n)}. \tag{1.1}$$

This ratio function can be used as a basis for hypothesis tests and confidence intervals by Wilks' theorem.

1.1.2 Empirical Likelihood Method

Empirical likelihood inference method is developed from the nonparametric likelihood method. For the ratio function in equation (1.1), suppose the arbitrary CDF F comes from the collection $\mathcal{F}_0$ of all CDF functions, then it is obvious that $0 \leq \mathcal{R}(F) \leq 1$, and $\mathcal{R}(F) = 0$ when F is continuous and $\mathcal{R}(F) = 1$ when $F = F_n$. Therefore, it might be trivial to consider such a large distribution set $\mathcal{F}_0$ and we may want to narrow the range of CDF F to a subset of $\mathcal{F}_0$ by excluding all continuous CDFs. This motivates the empirical likelihood.

Given the observations $x_1, x_2, \ldots, x_n$ and fixed sample size n, simply assume $p_i = \mathbb{P}(X = x_i),\ i = 1, 2, \ldots, n$, with $\sum_{i=1}^n p_i = 1$, and let $\mathcal{F}$ be the collection of all such distribution functions. For instance, suppose we are interested in the vector mean $\mu \in \mathbb{R}^d$ of i.i.d random vectors $X_1, X_2, \ldots, X_n$, by empirical likelihood method, we can write the ratio function (1.1) as

$$\mathcal{R}(\mu) = \sup\left\{\prod_{i=1}^n np_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i(X_i - \mu) = 0\right\}.$$

The following theorem, the so-called Wilks' theorem, says that the limiting distribution for the log of ratio function is a Chi-squared distribution, which is presented as the main result in Owen's book [1].

Theorem 1.1.1. *Let $X_1, X_2, \ldots, X_n$ be independent d-dimensional random vectors with* *common distributions F_0. Let $\mu_0 = \mathbb{E}(X_1) \in \mathbb{R}^d$ and suppose that $0 < \mathbb{E}(||X_1||^2) < \infty$. Then $-2\log(\mathcal{R}(\mu_0))$ converges in distribution to $\chi^2(d)$ as $n \to \infty$.*

Remark 1.1.1. *The above theorem has the following feature: the Chi-squared limit is the same as what we find in parametric likelihood models. Based on Theorem 1.1.1, we may use the Chi-squared limit to construct confidence regions or hypothesis tests just like what we do with the parametric likelihood method. That is, the confidence region is*

$$I_\gamma(\mu) = \left\{\mu \;\middle|\; -2\log(\mathcal{R}(\mu)) \leq \chi^2_\gamma(d)\right\}.$$

where γ is the confidence level and $\chi^2_\gamma(d)$ is the γ-quantile of Chi-squared distribution with degrees of freedom d.

The flexibility of the empirical likelihood method is that it can serve as a building block for many other extensions with wide applications of statistical inference. When applying Theorem 1.1.1, we may simply treat $X_1, X_2, \ldots, X_n$ as the input, the so-called *pseudo samples*, of the building block, without i.i.d assumption on $X_1, X_2, \ldots, X_n$. One such extension is to replace X_i with general estimating equations, which is developed in [2].

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d random variables with unknown distribution function F and θ is a p-dimensional parameter associated with F through a r-dimensional estimating vector function $g(x, \theta) = (g_1(x,\theta), g_2(x,\theta), \ldots, g_r(x,\theta))^T$ where $r \geq p$, such that $\mathbb{E}_F\{g(X_1, \theta)\} = 0$. A direct inference approach for θ may be hard to develop, but it is easy with some small changes in empirical likelihood method. We can use $g(X_i, \theta),\ i = 1, 2, \ldots, n$, as the pseudo samples of the building block as mentioned above. The empirical likelihood ratio function becomes

$$\mathcal{R}(\theta) = \sup\left\{\prod_{i=1}^n np_i \;\middle|\; p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g(X_i, \theta) = 0\right\}.$$

and [2] shows that a result similar to Theorem 1.1.1 holds with a Chi-squared limiting distribution with degrees of freedom p. Based on that result, interval estimation and hypothesis tests are easy to develop. It is widely believed that the key for this method is the restriction $\sum_{i=1}^{n} p_i g(X_i, \theta) = 0$ on the input of the empirical likelihood method, which makes the modeling more flexible in real world applications.

1.2 Copula and Tail Dependence

1.2.1 Copula

The study of copulas stems from question about the bivariate distribution function. Given two random variables X_1 ad X_2 with marginal distributions F_1 and F_2, what can be said about their joint distribution function F? Under the condition of independence between X_1 and X_2, it is trivial to see that $F(x_1, x_2) = F_1(x_1)F(x_2)$ so that the joint distribution is fully characterized by the two marginal distributions. Thus, the above question can be rephrased as: what can be said about the relationship between joint and marginal distributions under the dependence structure?

The milestone work on this question is by Sklar in his paper [3]. He introduced the concept *copula* and proved the well-known theorem that now bears his name, the Sklar's Theorem.

Definition 1.2.1. *A d-copula is a d-dimensional function C with the following properties:*

1. *The domain of C is $[0, 1]^d$.*
2. *C is grounded, i.e. $C(u_1, , u_{i1}, 0, u_{i+1}, , u_n) = 0$ for any $1 \leq i \leq n$ and $u_j \in [0, 1], j \neq i$.*
3. *C is d-increasing, i.e. for each hyperrectangle B in $[0, 1]^n$ the C-volume of B is non-negative.*
4. *For $u_i \in [0, 1]$, $1 \leq i \leq n$, $C(1, , 1, u_i, 1, , 1) = u_i$, where u_i is the i-th coordinate and all the other coordinates are 1.*

One important property of a d-copula C is that it is a *Lipschitz function*, namely, for $\mathbf{u} = (u_1, u_2, \ldots, u_d), \mathbf{v} = (v_1, v_2, \ldots, v_d) \in [0,1]^d$, one has

$$|C(\mathbf{u}) - C(\mathbf{v})| \leq \sum_{i=1}^{d} |u_i - v_i|.$$

Another important property is the *Fréchet-Hoeffding bounds* for a copula C. For $\mathbf{u} \in [0,1]^d$, one has

$$\max\{u_1 + u_2 + \ldots + u_d - d + 1, 0\} \leq C(\mathbf{u}) \leq \min\{u_1, u_2, \ldots, u_d\}.$$

Note that the upper bound is also a d-copula but the lower bound is a copula only when $d = 2$.

Sklar's Theorem tells the story of the existence of copula.

Theorem 1.2.1 (Sklar's Theorem)**.** *Let F be a joint distribution function with margins F_1 and F_2. Then there exists a copula C such that for all x, y in $\mathbb{R}$,*

$$F(x, y) = C(F_1(x), F_2(y)). \tag{1.2}$$

If F_1 and F_2 are continuous, then C is unique; otherwise, C is uniquely determined on Ran$(F_1) \times$ Ran(F_2) where Ran(F_i) is the range of F_i for $i = 1, 2$. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function F defined by (1.2) is a joint distribution function with margins F_1 and F_2.

This theorem states an interesting fact about the relationship among the joint distribution, the dependence structure, and the marginal distributions. The joint distribution, which includes all the information about dependence and margins, can be simply divided into two parts: the first part is the two marginal distributions, which includes all the information of margins but none of the dependence structure; the second part is the copula, which includes all the information of dependence structure but none of the margins.

Theorem 1.2.1 has a simple extension to d-dimensional joint distribution F. For more details about copulas, we refer to Nelsen's book [4].

Theorem 1.2.2 (Sklar's Theorem)**.** *Let F be a d-dimensional joint distribution function with margins $F_i,\ i = 1, 2, \ldots, d$. Then there exists a d-dimensional copula, or d-copula, C such that for all $x_1, x_2, \ldots, x_d$ in $\mathbb{R}$,*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)). \tag{1.3}$$

If F_i's are all continuous, then C is unique; otherwise, C is uniquely determined on $\prod_{i=1}^{d} Ran(F_i)$ where $Ran(F_i)$ is the range of F_i. Conversely, if C is a d-dimensional copula and F_i's are all distribution functions, then the function F defined by (1.3) is a d-dimensional joint distribution function with margins $F_i,\ i = 1, 2, \ldots, d$.

In many dependence modeling procedures, one may be interested in the lifetime of random variables and associated correlation. Then a direct modeling on the *survival copula* is more interesting. We use a 2-dimensional random vector as example. Suppose (X, Y) has marginal distributions F_1 and F_2 and copula C. We refer to $\widetilde{C}$ as the survival copula if for any real values $u, v \in [0, 1]$,

$$\widetilde{C}(u, v) = \mathbb{P}(1 - F_1(X) < u, 1 - F_2(Y) < v). \tag{1.4}$$

Survival copulas have two features. First, a survival copula is also a copula as it satisfies the copula Definition 1.2.1. Second, a survival copula $\widetilde{C}$ and its associated copula C have a transformation relationship: for all $u, v \in [0, 1]$,

$$\widetilde{C}(u, v) = u + v - 1 + C(u, v). \tag{1.5}$$

Therefore, survival copula can also be seen as the dependence structure of random variables, and modeling based on survival copulas sometimes may be more convenient and easier to interpret.

Statistical inference procedures for copulas lead to parametric, semi-parametric and non-parametric methods assuming i.i.d. observations with dependent components. For example, Joe [5] provides a great treatment of multivariate copula models as well as a review of maximum likelihood (ML) estimation of parameters in the models. [6] considers the problem of semiparametric inference for multivariate copula models with nonparametric marginal estimators and parametric estimators for copula densities. [7] and [8] propose a standard way to construct nonparametric estimators for copulas based on inversion formula of Sklar's Theorem. Such work has greatly explored the interdisciplinary applications of copula models in the fields of finance, insurance, risk management, econometrics and environmental science.

1.2.2 Association Measures

As mentioned before, copula is the dependence structure of a joint distribution. Applications of copulas are very popular in various fields including statistics, risk management, quantitative finance, econometrics and actuarial science. The name of copula stands for the dependence structure, a common sense in research and application.

Some association measures are developed based on copulas which uncover the main information about the dependence, like positive or negative dependence, and strength of the dependence. We introduce two basic and widely used association measures which are related to copulas.

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be independent vectors of continuous random variables, with joint distribution functions F, and common margins F_1 (of X_1 and X_2) and F_2 (of Y_1 and Y_2). Let C denote the copulas of (X_1, Y_1) and (X_2, Y_2) so that $F(x, y) = C(F_1(x), F_2(y))$. Then

- *Kendall's tau*:

$$\begin{aligned}\tau &= \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) \\ &= 4 \iint_{[0,1]^2} C(u,v) dC(u,v) - 1.\end{aligned}$$

- *Spearman's rho*:

$$\begin{aligned}\rho &= 3 \left\{\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)\right\} \\ &= 12 \iint_{[0,1]^2} C(u,v) du dv - 3.\end{aligned}$$

These two association measures are single values between -1 and 1. Their positiveness and negativeness show the positive and negative correlation respectively, and their absolute values show the strength of the correlations. One advantage of association measures over copulas is that the association measures are simple values which for certain specific information are easy to measure, understand and interpret. On the other hand, since copulas are functions, they might not provide simple and clear dependence information at a first glance. Another advantage is that the association measures are defined only on copulas with no information of margins included. So they are insensitive to marginal transformation which is admired by practitioners. Therefore, these association measures are preferred in real applications over copulas, but the inference approaches are still developed based on their representation with copulas since in nature the information provided by these measures are all derived from the dependence structure, i.e. copulas.

Other association measures are also available in the literature. We refer to Chapter 5 of [4] for more details.

1.2.3 Tail Dependence

Kendall's tau and Spearman's rho are among many of the dependence concepts which are designed to describe how large (or small) values of one random variable appear with large (or small) values of the other over the entire domain. However, when it comes to extreme events or correlations, this type of dependence is not suitable. Another concept is *tail dependence*, which measures the dependence between the variables on the upper-right (or low-left) tail regions. The concept of tail dependence is closely related to the *multivariate extreme value theory*, since both take the limiting distributions on tail regions into account in the modeling procedure.

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent and identically distributed random vectors with distribution function F and marginal distributions F_1 and F_2, i.e. $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. Bivariate Extreme Value Theory assumes that there are constants $a_n > 0, c_n > 0, b_n \in R, d_n \in R$ such that

$$\lim_{n\to\infty} \mathbb{P}\Big(a_n\Big(\max_{1\le i\le n} X_i - b_n\Big) \le x, c_n\Big(\max_{1\le i\le n} Y_i - d_n\Big) \le y\Big) = G(x, y), \tag{1.6}$$

for all continuous points (x, y) of G. In this case, G is called an *extreme value distribution* and F is said to belong to the *domain of attraction of* G.

One important concept is the *tail dependence function*. It follows from (1.6) that the following dependence convergence holds:

$$l(x, y) = \lim_{t\to 0} t^{-1}\Big\{1 - F\big((1 - F_1)^-(tx), (1 - F_2)^-(ty)\big)\Big\} \tag{1.7}$$

for all $x, y \ge 0$, where $G_1(x) = G(x, \infty)$, $G_2(y) = G(\infty, y)$, and $(\cdot)^-$ denotes the left continuous inverse function. We refer to [9] for more details. It is easy to check that $l(ax, ay) = al(x, y)$ for all $a, x, y \ge 0$ and $x \vee y \le l(x, y) \le x + y$. This homogeneous property has been employed to extrapolate data into a tail region so that extreme events can be predicted (for details, see for example, [10]). However, when $l(x, y) = x + y$, equation (1.7) implies that

$$\lim_{t\to 0} t^{-1}\mathbb{P}(1 - F_1(X_1) < tx, 1 - F_2(Y_1) < ty) = 0, \tag{1.8}$$

which makes extrapolation, i.e. statistical inference, impossible for concomitant extreme sets. In this case, F is said to have the asymptotic independence property, and a different convergence rate condition in (1.8) is needed for predicting joint extreme events. In other words, extreme value condition (1.6) is not enough for predicting extreme events in case of asymptotic independence. If the limit in (1.8) is not identical to zero, then F is said to have the *asymptotic dependence* property. It is known that a bivariate normal distribution with correlation coefficient between -1 and 1 is asymptotically independent, i.e. (1.8) holds (for details, see [11]).

Another important concept is *upper tail dependence coefficient*. The upper tail dependence coefficient λ_U is the limit (if exists) of the conditional probability that Y_1 is greater than the $100t$-th percentile of F_2 given that X_1 is greater than the $100t$-th percentile of F_1 as t approaches 1, i.e.

$$\lambda_U = \lim_{t \downarrow 0^+} \mathbb{P}(1 - F_2(Y_1) \leq t | 1 - F_1(X_1) \leq t) \tag{1.9}$$

The tail dependence coefficient measures the extreme events on tail regions without giving any consideration to marginal distributions, which is similar to copula. In fact, the tail dependence coefficient has a representation in terms of survival copula defined in equation (1.4)

$$\lambda_U = \lim_{t \downarrow 0^+} t^{-1} \widetilde{C}(t, t). \tag{1.10}$$

Equation (1.10) provides a way to apply inference methods to the tail dependence coefficient via the estimation of survival copula. However, to achieve good statistical properties, conditions in extreme value theory are necessary since this limit indicates the tail dependence coefficient is defined based on the limiting distribution instead of the distribution of the observations. Therefore, it is necessary to control the rate of convergence between observations and their limit during the modeling procedure.

1.3 Risk Measures

In this section, we introduce the concepts of two simple risk measures in risk management, the Value-at-Risk (V@R) and the Expected Shortfall (ES). These two risk measures play an important role in modern financial and insurance markets. They are well studied in the literature, and inference methods have been widely proposed as well. In our research, these two risk measures are fundamental to other complicated risk measures.

Value-at-Risk is used to quantify the value of assets' tail risk. It represents the potential maximal loss one may have during a given period with a given risk level (probability). Compared to variance, it provides a more sensible measure of the risk of the assets since it focuses on losses.

The definition of V@R is related to the quantile of a random variable in statistics.

Definition 1.3.1. *Suppose X is a random variable with continuous distribution function F. Given the risk level $\alpha \in (0,1)$, the Value-at-Risk of X at risk level α is*

$$V@R_\alpha(X) = \inf\left\{x\middle|\mathbb{P}(X \le x) \ge \alpha\right\} = F^{-}(\alpha). \tag{1.11}$$

where F^{-} is the inverse function of F.

The Expected Shortfall is the average value of the loss that exceeds the V@R. The point of this risk measure is to have a metric close to the V@R but that takes into account the maximal losses exceeding a given risk level.

Definition 1.3.2. *Suppose X is a random variable with continuous distribution function F. Given the risk level $\alpha \in (0,1)$, the Expected Shortfall of X at risk level α is*

$$ES_\alpha(X) = \mathbb{E}\{X|X > V@R_\alpha(X)\} = \frac{1}{1-\alpha}\int_\alpha^1 V@R_\gamma(X)d\gamma. \tag{1.12}$$

Inference methods for V@R and ES has been also well studied in the literature. For V@R, a sample quantile estimation is an intuitive way to construct a point estimator and its asymptotic result is available in statistics theory so the interval estimation and tests are easy to develop. The drawback is that when the risk level is high, the inference approach requires extremely large sample size to achieve normal stability. Another approach is Monte Carlo simulation which is frequently used in practice to estimate V@R; see [12]. To effectively construct interval estimation of V@R for regulatory purpose, empirical likelihood method for a quantile can be employed; see [13]. For ES, [14] develops nonparametric estimation approach under dependence condition. [15] proposes the empirical likelihood method for the interval estimation. Their work leads to inference methods of V@R and ES, contributing to the ease of implementation for practical applications.

CHAPTER 2

INFERENCE FOR CONDITIONAL VALUE-AT-RISK

Value-at-risk is one commonly employed risk measure for tail risk in financial markets. Based on a predictive regression, [16] proposed inferences for conditional risk measures. This Chapter first corrects the wrong formula of the asymptotic variance of the conditional Value-at-Risk estimator via the least squares estimate in [16]. It turns out that the asymptotic variance depends on whether the model has a predicting variable with an infinite variance, which makes it quite challenging in quantifying the uncertainty of the conditional risk estimator. Therefore this chapter further proposes a unified empirical likelihood method for constructing a confidence interval for the conditional Value-at-Risk regardless of whether there exists an infinite variance predicting variable in the predictive regression. The content of this chapter is based on Y. He, Y. Hou and L.Peng (2016). Inference for Conditional Value-at-Risk of a Predictive Regression. *Preprint*.

2.1 Introduction

This chapter is mainly motivated by the study in [16] on inferring a conditional Value-at-Risk (V@R) based on a predictive regression model. After giving a rigorous derivation of the asymptotic limit of the above conditional V@R estimator and stating the correct asymptotic variance, which has a quite complicated formula and depends on whether some predicting variables have an infinite variance, we propose an effective empirical likelihood method to construct an interval for the conditional V@R. The proposed unified method does not need to estimate the complicated asymptotic variance and works regardless of predicting variables having a finite variance or an infinite variance. Therefore, the new empirical likelihood inference is quite applicable. A more detailed introduction goes as follows.

An important aspect in risk management is to infer a risk measure with accurate statistical uncertainty quantification. Some well-known risk measures include coherent risk measures (see [17, 18, 19, 20]), distortion risk measures (see [21]), Wang's premium principle and proportional hazard transform risk measures (see [22, 23, 24, 25, 26, 27]), and Haezendonck-Goovaerts risk measure in actuarial science (see [28, 29]). And it is argued that V@R is one of most commonly employed risk measures in financial and insurance industries and is much preferred in terms of measuring economic tail risk; see, e.g., [30, 31, 32, 33, 34].

The V@R of a random variable Y at level $\alpha \in (0,1)$ is given by

$$V@R(\alpha, Y) = \inf\{q : F_Y(q) \geq \alpha\}, \quad \text{i.e.,} \quad V@R(\alpha, Y) = F_Y^{-1}(\alpha),$$

where F_Y is the cumulative distribution function of Y and $F_Y^{-1}(y)$ is the inverse function of $F_Y(y)$. Given identically distributed observations of sample size n, say, $Y_1, \ldots, Y_n$, the V@R at level α can be estimated non-parametrically by the $[n\alpha]$-th smallest observation, whose asymptotic variance depends on the density of the underlying distribution. Alternatively, often with substantial computational cost, Monte Carlo simulation is also frequently used to estimate V@R; see, e.g., [35]. For evaluating different V@R forecast models one may refer to, e.g., [36] and [37].

To effectively construct interval estimation of a V@R for regulation purpose, empirical likelihood method for a quantile can be employed; see [38] and [39]. Empirical likelihood is a distribution-free statistical inference method based on a data-driven likelihood ratio function. As proposed by [38], with a smooth distribution function K and a bandwidth parameter h, the empirical likelihood ratio function for $V@R(\alpha, Y)$ may be taken as

$$L(q) = \sup\left\{\Pi_{t=1}^{n}(np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t K\left(\frac{q - Y_t}{h}\right) = \alpha\right\}, \quad q \in \mathbb{R}.$$

The resulting asymptotic confidence interval with level γ is of the form $\{q \in \mathbb{R} : -2\log L(q) \leq$

$\chi^2_{1,\gamma}\}$, where $\chi^2_{1,\gamma}$ is the γ-quantile of a chi-squared distribution with one degree of freedom since Wilks theorem holds for the above empirical likelihood method under some weak conditions, i.e., $-2\log L(V@R(\alpha, Y))$ is asymptotically chi-squared distributed with one degree of freedom as $n \to \infty$. The above interval can be effectively determined by using a standard search algorithm and often has good finite-sample coverage accuracy and some good theoretical properties such as Bartlett correctablity. We refer to [1] for an overview on empirical likelihood method, which has been proved to be quite effective in interval estimation and hypothesis tests.

When an asset or a financial variable can be predicted by some market variables or risk factors, a conditional risk measure given the market situation would be more meaningful than an unconditional risk measure. In particular, a conditional V@R is defined as a conditional quantile given the market situation. In this case, a model-free estimator of the conditional V@R can be constructed by kernel smoothing techniques, which results in a slower rate of convergence than the standard rate $n^{-1/2}$; see, e.g., [40]. Alternatively one could directly model the conditional quantile as a parametric form of the market variables and so quantile regression techniques can be employed to infer the conditional V@R, which gives the standard rate of convergence $n^{-1/2}$; see [41]. Here we revisit the study in [16] on estimating a conditional V@R based on a predictive regression model.

Suppose that we have n observations (data points) $\{Y_t, \boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,k})^T\}_{t=1}^n$ from the following simple predictive regression:

$$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{t,i} + \epsilon_t, \quad t = 1, \ldots, n, \tag{2.1}$$

where $\{\epsilon_t\}_{t=1}^n$ is a sequence of independent and identically distributed random variables with zero mean and finite second-order moment and independent of the predicting variables $\{\boldsymbol{X_t}\}_{t=1}^n$, which could be a stationary sequence rather than an independent sequence. Based on (2.1), the conditional V@R at level $\alpha \in (0,1)$ of Y_t given $\boldsymbol{X}_t = \boldsymbol{x} := (x_1, \cdots, x_k)^T$ is defined as

$$V@R_{\boldsymbol{x}}(\alpha) := \inf\{q : P(Y_t \le q | \boldsymbol{X}_t = \boldsymbol{x}) \ge \alpha\} = F_\epsilon^{-1}(\alpha) + \boldsymbol{z}^T\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $\boldsymbol{z} = (1, \boldsymbol{x}^T)^T$ and F_ϵ denotes the distribution function of ϵ_t. Therefore a simple nonparametric estimator for the above conditional V@R is

$$\widehat{V@R}_{\boldsymbol{x}}(\alpha) = \widehat{\epsilon}_{n,[n\alpha]} + \boldsymbol{z}^T\widehat{\boldsymbol{\beta}},$$

where $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, $\widehat{\epsilon}_t = Y_t - \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_t$ with $\boldsymbol{Z}_t = (1, \boldsymbol{X}_t^T)^T$, $\widehat{\epsilon}_{n,1} \le \cdots \le \widehat{\epsilon}_{n,n}$ denote the order statistics of $\widehat{\epsilon}_1, \cdots, \widehat{\epsilon}_n$.

When $\widehat{\boldsymbol{\beta}} = \{\frac{1}{n}\sum_{t=1}^n \boldsymbol{Z}_t\boldsymbol{Z}_t^T\}^{-1}\frac{1}{n}\sum_{t=1}^n Y_t\boldsymbol{Z}_t$, i.e., the least squares estimate, [16] derived the asymptotic limit of $\widehat{V@R}_{\boldsymbol{x}}(\alpha)$ when $\boldsymbol{X}_t's$ are independent and identically distributed random vectors without mentioning whether finite (second-order) moment conditions are needed.

In this chapter, we first notice that the asymptotic variance of $\widehat{V@R}_{\boldsymbol{x}}(\alpha)$ derived in [16] is miscalculated even when $E(\boldsymbol{X}_t\boldsymbol{X}_t^T) < \infty$ is assumed; see Remark 2.2.1 below. After providing a correct formula with rigorous proofs, which depends on whether all predicting variables have a finite variance or some predicting variables have an infinite variance, we find that the asymptotic variance is too complicated and so interval estimation and hypothesis test for the conditional V@R become nontrivial at all. Therefore we further propose a smooth empirical likelihood method to effectively construct a confidence interval for the conditional V@R without estimating the asymptotic variance explicitly and knowing whether all predicting variables have a finite variance or some of predicting variables have an infinite variance. This avoids the challenging question of testing whether a variable has a finite variance or an infinite variance. Therefore the proposed confidence interval for the conditional V@R given the current market situation becomes practically useful in monitoring the risk of the underlying asset or financial variable.

We organize this chapter as follows. The asymptotic limit of the conditional V@R estimator based on the least squares estimator is given in Section 2.2. Section 2.3 discusses the empirical likelihood method for unified interval estimation. A simulation study and data analysis are presented in Sections 2.4 and 2.5, respectively. Some conclusions are summarized in Section 2.6.

2.2 Asymptotic Results

In this section we shall correct the wrong asymptotic variance formula derived in [16] by considering the cases of finite variance and infinite variance separately, and by allowing $\{\boldsymbol{X}_t\}$ to be a stationary sequence, which is important for financial applications.

First we consider the case that all predicting variables in the regression model (2.1) have a finite variance. More specifically we assume the following conditions:

- A) Assume $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean, finite variance σ^2 and is independent of $\{\boldsymbol{X}_t\}$, which is a stationary sequence with finite variances and satisfies the weak law of large numbers: $\frac{1}{n}\sum_{t=1}^n \boldsymbol{X}_t \xrightarrow{P} E\boldsymbol{X}_1$ and $\frac{1}{n}\sum_{t=1}^n \boldsymbol{X}_t\boldsymbol{X}_t^T \xrightarrow{P} E(\boldsymbol{X}_1\boldsymbol{X}_1^T)$ as $n \to \infty$, where $\xrightarrow{P}$ denotes the convergence in probability. Further assume $\Omega := E(\boldsymbol{Z}_1\boldsymbol{Z}_1^T)$ is positive definite, $E(\|\boldsymbol{X}_t\|^{2+\delta_0}) < \infty$ for some $\delta_0 > 0$ and $F_\epsilon'(y)$ is continuous and positive at $y = F_\epsilon^{-1}(\alpha)$.

Theorem 2.2.1. *Under conditions A) above, as $n \to \infty$,*

$$\sqrt{n}\{\widehat{V@R}_{\boldsymbol{x}}(\alpha) - V@R_{\boldsymbol{x}}(\alpha)\} \xrightarrow{d} N(0, \omega^2 + \sigma^2\boldsymbol{z}^T\Omega^{-1}\boldsymbol{z} + \Delta),$$

where $\boldsymbol{z} = (1, \boldsymbol{x}^T)^T$, $\omega^2 = \frac{\alpha(1-\alpha)}{\{F'_\epsilon(F_\epsilon^{-1}(\alpha))\}^2}$, *and* $\Delta = \Delta_1 + \Delta_2$ *with*

$$\begin{aligned}\Delta_1 =&\sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1) + 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1) \quad \textit{and}\\ \Delta_2 =&-2\sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z} - 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z}.\end{aligned}$$

Remark 2.2.1. *The last term Δ in the above asymptotic variance is overlooked in [16]. Its first part Δ_1 is due to the impact of the statistical uncertainty of the least squares estimator $\widehat{\boldsymbol{\beta}}$ on residual quantile estimation. The second part Δ_2 attributes to the interaction between $\widehat{\boldsymbol{\beta}}$ and the residual quantile estimation, and it is easy to verify that $\Delta_2 = 0$ if ϵ_t is normally distributed.*

Next we consider the case when some of predicting variables have an infinite variance. Without loss of generality we only consider $EX_{t,k}^2 = \infty$ since the case with more than one infinite-variance predictors can be shown similarly. In this case, we assume the following conditions:

- B) Assume $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance σ^2 and is independent of $\{\boldsymbol{X}_t\}$, which is a stationary sequence with $EX_{t,i}^2 < \infty$ for $i = 1, \cdots, k-1$ and the distribution function F_k of $X_{t,k}$ lies in the domain of attraction of a stable law with index $d \in (1, 2)$ (see Feller (1971) for details on stable law). Put $\widetilde{\boldsymbol{X}}_t = (X_{t,1}, \cdots, X_{t,k-1})^T$ and assume $\frac{1}{n}\sum_{t=1}^n \boldsymbol{X}_t \xrightarrow{P} E\boldsymbol{X}_1$, $\frac{1}{n}\sum_{t=1}^n \widetilde{\boldsymbol{X}}_t\widetilde{\boldsymbol{X}}_t^T \xrightarrow{P} E(\widetilde{\boldsymbol{X}}_1\widetilde{\boldsymbol{X}}_1^T)$ and $\frac{\sum_{t=1}^n (X_{t,k} - E(X_{t,k}))}{n^{1/d}L(n)}$ converges in distribution to a stable law with index d, where $L(n)$ is a slowly varying function, i.e., $L(tx)/L(t) \to 1$ for any $x > 0$ as $t \to \infty$. Further assume $\widetilde{\Omega} := E(\widetilde{\boldsymbol{Z}}_1\widetilde{\boldsymbol{Z}}_1^T)$ is positive definite, $E(\|\widetilde{\boldsymbol{X}}_t\|^{2+\delta_0}) < \infty$ for some $\delta_0 > 0$, and $F'_\epsilon(y)$ is continuous and positive at $y = F_\epsilon^{-1}(\alpha)$, where $\widetilde{\boldsymbol{Z}}_t = (1, X_{t,1}, \cdots, X_{t,k-1})^T$.

Theorem 2.2.2. *Under conditions B) above, as $n \to \infty$*

$$\sqrt{n}\{\widehat{V@R}_{\boldsymbol{x}}(\alpha) - V@R_{\boldsymbol{x}}(\alpha)\} \xrightarrow{d} N(0, \omega^2 + \sigma^2\widetilde{\boldsymbol{z}}^T\widetilde{\Omega}^{-1}\widetilde{\boldsymbol{z}} + \widetilde{\Delta}),$$

where $\widetilde{\boldsymbol{z}} = (1, x_1, \cdots, x_{k-1})^T$, $\omega^2 = \frac{\alpha(1-\alpha)}{\{F'_\epsilon(F_\epsilon^{-1}(\alpha))\}^2}$, *and* $\widetilde{\Delta} = \widetilde{\Delta}_1 + \widetilde{\Delta}_2$ *with*

$$\begin{aligned}\widetilde{\Delta}_1 =& \sigma^2 E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}E(\widetilde{\boldsymbol{Z}}_1) + 2\frac{E(\epsilon_1 I(\epsilon_1 \leq F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}E(\widetilde{\boldsymbol{Z}}_1) \quad \textit{and} \\ \widetilde{\Delta}_2 =& -2\sigma^2 E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}\widetilde{\boldsymbol{z}} - 2\frac{E(\epsilon_1 I(\epsilon_1 \leq F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}\widetilde{\boldsymbol{z}}.\end{aligned}$$

Remark 2.2.2. *Theorem 2.2.2 is very similar to Theorem 2.2.1, except that the infinite-variance predictor does not play a role in the asymptotic limit since* $\widehat{\beta}_k - \beta_k$ *has a faster rate of convergence than other parts. Again, the above asymptotic variance is different from that given in [16], where moment conditions on predicting variables are not stated at all.*

2.3 Empirical Likelihood Method

It is known that quantifying the uncertainty of a risk measure is important and challenging in risk management. Since the asymptotic variance of the above nonparametric estimator of the conditional V@R is quite complicated, directly estimating the asymptotic variance is nontrivial at all especially because the asymptotic variance depends on whether there exist some infinite-variance predictor(s). Moreover, the applicability of a bootstrap method remains unknown due to dependent and possibly infinite variance predicting variables. Here we propose the following empirical likelihood method based on a combination of estimating equations in [2] and the smoothing technique in [38]. Write $\boldsymbol{Z}_t = (1, X_{t,1}, \cdots, X_{t,k})^T$ for $t \geq 1$ and put

$$\boldsymbol{W}_t(\boldsymbol{\beta}, \theta) =: \frac{1}{\|\boldsymbol{Z}_t\|^2}\widetilde{\boldsymbol{W}}_t(\boldsymbol{\beta}, \theta) =: \frac{1}{\|\boldsymbol{Z}_t\|^2}\left(\widetilde{W}_{t,1}(\boldsymbol{\beta}, \theta), \cdots, \widetilde{W}_{t,k+2}(\boldsymbol{\beta}, \theta)\right)^T$$

with

$$\widetilde{W}_{t,1}(\boldsymbol{\beta},\theta) = K(\frac{\theta - \boldsymbol{\beta}^T\boldsymbol{z} - (Y_t - \boldsymbol{\beta}^T\boldsymbol{Z}_t)}{h}) - \alpha,$$
$$\widetilde{W}_{t,i+2}(\boldsymbol{\beta}) = (Y_t - \boldsymbol{\beta}^T\boldsymbol{Z}_t)X_{t,i} \quad \text{for} \quad i = 0, 1, \ldots, k,$$

where K is a smooth distribution function and $h = h(n) > 0$ is a bandwidth, and for presentation convenience $X_{t,0} := 1$ for all $t \geq 1$. The reason to employ the weight $\|\boldsymbol{Z}_t\|^{-2}$ is to get rid of the effect of infinite moments of $\boldsymbol{X}_t$ so as to have a unified inference procedure. One can choose different weight functions such as $\max\limits_{1\leq i\leq k+1} Z_{t,i}^2$. The observations $\widetilde{\boldsymbol{W}}_t(\boldsymbol{\beta},\theta)$ can be less under-weighted (or even not weighted at all), if the first (and second) moment(s) of $\|\boldsymbol{Z}_t\|$ is(are) finite; see Remarks 2.3.1 and 2.3.2 below. Choosing an optimal weight function is beyond the scope of this chapter. Next we define the empirical likelihood function for $\boldsymbol{\beta}$ and θ as

$$L(\boldsymbol{\beta},\theta) = \sup\left\{\prod_{t=1}^{n}(np_t) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \boldsymbol{W}_t(\boldsymbol{\beta},\theta) = 0\right\}.$$

Since we are only interested in $\theta^0 = V@R_{\boldsymbol{x}}(\alpha)$, we consider the so-called profile empirical likelihood function $L^P(\theta) = \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta},\theta)$.

To establish a Wilks type of result we assume the following conditions:

- C1) $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and $E(|\epsilon_t|^{2+\delta_0})$ for some $\delta_0 > 0$ and it is independent of the stationary sequence $\{\boldsymbol{X}_t\}$, which satisfies the weak law of large numbers:

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{Z}_t}{\|\boldsymbol{Z}_t\|^2} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1}{\|\boldsymbol{Z}_1\|^2}\right), \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^2} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1\boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^2}\right)$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^4} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1\boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^4}\right);$$

Furthermore, assume $F'_\epsilon(y)$ is continuous and positive in a neighborhood of $y = F_\epsilon^{-1}(\alpha)$.

- C2) Define

$$\Sigma_1 := E\left\{ \begin{pmatrix} \frac{I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)) - \alpha}{\|\boldsymbol{Z}_1\|^2} \\ \epsilon_1 \boldsymbol{Z}_1 / \|\boldsymbol{Z}_1\|^2 \end{pmatrix} \left(\frac{I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)) - \alpha}{\|\boldsymbol{Z}_1\|^2}, \quad \epsilon_1 \boldsymbol{Z}_1^T / \|\boldsymbol{Z}_1\|^2 \right) \right\},$$

$$\Sigma_2 := \begin{pmatrix} F'_\epsilon(F_\epsilon^{-1}(\alpha)) E\left((\boldsymbol{Z}_1 - \boldsymbol{z})^T / \|\boldsymbol{Z}_1\|^2 \right) \\ -E\left(\boldsymbol{Z}_1 \boldsymbol{Z}_1^T / \|\boldsymbol{Z}_1\|^2 \right) \end{pmatrix},$$

and assume Σ_1 is positive definite.

- C3) Assume $K(x) = \int_{-1}^{x} g(s)\, ds$, where $g(x)$ is a symmetric density function on $[-1, 1]$ with bounded derivative. Further assume $nh^4 \to 0$ and $nh^{r_0} \to \infty$ for some $r_0 \in (2, 4)$.

Theorem 2.3.1. *Under conditions C1)–C3) above, $-2 \log L^P(V@R_{\boldsymbol{x}}(\alpha))$ converges in distribution to a chi-squared limit with one degree of freedom as $n \to \infty$.*

Based on Theorem 2.3.1, an asymptotic confidence interval for $V@R_{\boldsymbol{x}}(\alpha)$ with level $\xi \in (0, 1)$ is

$$I_\xi = \{\theta : -2 \log L^P(\theta) \le \chi^2_{1,\xi}\},$$

where $\chi^2_{1,\xi}$ denotes the ξ-quantile of $\chi^2(1)$. Like Qin and Lawless (1994), we could study the asymptotic limit of the maximum empirical likelihood estimator defined by $\widehat{\theta}_{MELE} = \arg\max_\theta L^P(\theta)$, which has a different asymptotic variance from that given in Theorems 2.2.1 and 2.2.2 since the proposed profile empirical likelihood method is based on a weighted least squares estimate instead of the ordinary least squares estimate.

Remark 2.3.1 (Predictors with finite mean)**.** *If we replace the above $\boldsymbol{W}_t$ by $\|\boldsymbol{Z}_t\|^{-1} \widetilde{\boldsymbol{W}}_t$, and replace the weights $\|\boldsymbol{Z}_t\|^{-2}$ by $\|\boldsymbol{Z}_t\|^{-1}$ (and $\|\boldsymbol{Z}_t\|^{-4}$ by $\|\boldsymbol{Z}_t\|^{-2}$) in conditions C1-C3 when $E(\|\boldsymbol{X}_t\|) < \infty$, then Theorem 2.3.1 still holds.*

Remark 2.3.2 (Predictors with finite variance)**.** *If we replace the above* $\boldsymbol{W}_t$ *by* $\widetilde{\boldsymbol{W}}_t$*, and neglect the weights* $\|\boldsymbol{Z}_t\|^{-2}$ *(and* $\|\boldsymbol{Z}_t\|^{-4}$*) in conditions C1-C3 when* $E(\|\boldsymbol{X}_t\|^2) < \infty$*, then Theorem 2.3.1 still holds.*

Remark 2.3.3 (GARCH errors)**.** *The proofs for Theorems 1–3 are still valid when independent* $\epsilon_t' s$ *are replaced by a GARCH(p,q) process.*

2.4 Simulation Study

In this section, we carry out a simulation study to illustrate the performance of the proposed unified empirical likelihood method. We choose various types of predictors and error distributions in the following predictive linear model:

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \epsilon_t. \tag{2.2}$$

We take $\beta_0 = 1, \beta_1 = 2, \beta_2 = 2$ in all simulated models, and select predictors $(X_{t,1}, X_{t,2})^T$ as combinations of independent student's t with 1.5 degrees of freedom ($t(1.5)$), stationary autoregressive process with order 1 (AR(1)) and stationary Garch model (Garch(1,1)). Error distribution (ϵ_t) is generated from either the standard normal (N(0,1)) or a centered log normal with parameter 0 and $1/16$ (LN(0,1/16)), i.e., the mean of the log normal is zero. Note that independent t(1.5) for $X_{t,i}$ has an infinite variance, the AR(1) sequence for $X_{t,i}$ is specified as

$$X_{t,i} = 0.355 X_{t-1,i} + \phi_t, \tag{2.3}$$

and a Garch(1,1) sequence for $X_{t,i}$ is given by

$$X_{t,i} = \sigma_t \eta_t, \quad \sigma_t^2 = 0.1 + 0.7 X_{t-1,i}^2 + 0.1 \sigma_{t-1}^2. \tag{2.4}$$

where $\phi_t' s$ and $\eta_t' s$ are independent and identically distributed standard normal random variables.

Consider the following parameter settings: sample size $n = 2000$ and 5000 for $\alpha = 0.95, 0.99$, the biweight kernel K, bandwidth $h = 0.5 \times n^{-1/3}$ or $1 \times n^{-1/3}$ or $1.5 \times n^{-1/3}$ and the conditional predictors $\mathbf{x} = (0.1, 0.1)^T$. Note that the order $n^{-1/3}$ for the bandwidth comes from the optimal bandwidth in smoothing distribution function estimation. For each setting, we repeat the calculation of the profile empirical likelihood function for 10000 times and compute the coverage probability of the proposed empirical likelihood method for the construction of 95% and 90% confidence intervals of $V@R_{\mathbf{x}}(\alpha)$. Besides, we also present the simulation results of the normal approximation intervals based on the wrong formula of the asymptotic variance in [16] denoted by $I_*(0.95)$ and $I_*(0.90)$ in Tables 2.1 – 2.4. Note that we use the true density F'_ϵ rather than kernel density estimation in estimating the wrong asymptotic variance in [16].

From Tables 2.1 – 2.4, we observe that i) the coverage probabilities for $I_*(0.95)$ and $I_*(0.90)$ well indicate that the asymptotic variance in [16] is incorrect; ii) results for the proposed unified empirical likelihood method show that the performance of the new method improves as the sample size increases, and the infinite variance has an effect on the coverage accuracy for a smaller sample size n and/or a large level α; iii) a larger sample size seems necessary for achieving an accurate interval for the considered high level $\alpha = 0.99$; iv) the new method is robust to the considered three choices of bandwidth h.

Table 2.1: Empirical coverage probabilities of the proposed unified empirical likelihood interval for the conditional $V@R_{\boldsymbol{x}}(\alpha)$ with error N(0,1), $h_i = 0.5i \times n^{-1/3}$ for $i = 1, 2, 3$, $\alpha = 0.95$ and $\boldsymbol{x} = (0.1, 0.1)^T$.

Model	t(1.5)+AR(1)		t(1.5)+Garch(1,1)		AR(1)+Garch(1,1)	
n	2000	5000	2000	5000	2000	5000
$I_*(0.95)$	0.9646	0.9679	0.9689	0.9703	0.9675	0.9696
$I_{h_1}(0.95)$	0.9492	0.9509	0.9497	0.9513	0.9513	0.9490
$I_{h_2}(0.95)$	0.9498	0.9507	0.9499	0.9506	0.9500	0.9493
$I_{h_3}(0.95)$	0.9491	0.9508	0.9498	0.9494	0.9508	0.9481
$I_*(0.90)$	0.9303	0.9291	0.9319	0.9314	0.9269	0.9283
$I_{h_1}(0.90)$	0.9047	0.8983	0.9023	0.9012	0.8999	0.8992
$I_{h_2}(0.90)$	0.9040	0.8978	0.9006	0.9006	0.8992	0.8990
$I_{h_3}(0.90)$	0.9025	0.8992	0.9002	0.9007	0.8996	0.8975

Table 2.2: Empirical coverage probabilities of the proposed unified empirical likelihood interval for the conditional $V@R_{\boldsymbol{x}}(\alpha)$ with error N(0,1), $h_i = 0.5i \times n^{-1/3}$ for $i = 1, 2, 3$, $\alpha = 0.99$ and $\boldsymbol{x} = (0.1, 0.1)^T$.

Model	t(1.5)+AR(1)		t(1.5)+Garch(1,1)		AR(1)+Garch(1,1)	
n	2000	5000	2000	5000	2000	5000
$I_*(0.95)$	0.9338	0.9469	0.9389	0.9462	0.9322	0.9506
$I_{h_1}(0.95)$	0.9715	0.9478	0.9705	0.9516	0.9654	0.9528
$I_{h_2}(0.95)$	0.9708	0.9470	0.9699	0.9510	0.9643	0.9527
$I_{h_3}(0.95)$	0.9692	0.9470	0.9693	0.9511	0.9648	0.9529
$I_*(0.90)$	0.8990	0.9031	0.9011	0.9052	0.8953	0.9101
$I_{h_1}(0.90)$	0.9191	0.8960	0.9141	0.9024	0.9003	0.9030
$I_{h_2}(0.90)$	0.9193	0.8954	0.9153	0.9041	0.9000	0.9013
$I_{h_3}(0.90)$	0.9196	0.8945	0.9180	0.9038	0.9038	0.9015

Table 2.3: Empirical coverage probabilities of the proposed unified empirical likelihood interval for the conditional $V@R_{\boldsymbol{x}}(\alpha)$ with error LN(0,1/16), $h_i = 0.5i \times n^{-1/3}$ for $i = 1, 2, 3$, $\alpha = 0.95$ and $\boldsymbol{x} = (0.1, 0.1)^T$.

Model	t(1.5)+AR(1)		t(1.5)+Garch(1,1)		AR(1)+Garch(1,1)	
n	2000	5000	2000	5000	2000	5000
$I_*(0.95)$	0.9978	1.0000	0.9984	1.0000	0.9984	1.0000
$I_{h_1}(0.95)$	0.9490	0.9511	0.9482	0.9497	0.9497	0.9489
$I_{h_2}(0.95)$	0.9489	0.9486	0.9473	0.9489	0.9465	0.9459
$I_{h_3}(0.95)$	0.9400	0.9415	0.9346	0.9388	0.9320	0.9359
$I_*(0.90)$	0.9906	1.0000	0.9933	1.0000	0.9927	1.0000
$I_{h_1}(0.90)$	0.9025	0.8985	0.8996	0.8999	0.8985	0.8968
$I_{h_2}(0.90)$	0.8980	0.8944	0.8961	0.8982	0.8962	0.8925
$I_{h_3}(0.90)$	0.8821	0.8878	0.8816	0.8818	0.8775	0.8747

Table 2.4: Empirical coverage probabilities of the proposed unified empirical likelihood interval for the conditional $V@R_{\boldsymbol{x}}(\alpha)$ with error LN(0,1/16), $h_i = 0.5i \times n^{-1/3}$ for $i = 1, 2, 3$, $\alpha = 0.99$ and $\boldsymbol{x} = (0.1, 0.1)^T$.

Model	t(1.5)+AR(1)		t(1.5)+Garch(1,1)		AR(1)+Garch(1,1)	
n	2000	5000	2000	5000	2000	5000
$I_*(0.95)$	0.9630	0.9729	0.9650	0.9728	0.9621	0.9740
$I_{h_1}(0.95)$	0.9469	0.9469	0.9506	0.9508	0.9469	0.9525
$I_{h_2}(0.95)$	0.9439	0.9458	0.9511	0.9500	0.9448	0.9513
$I_{h_3}(0.95)$	0.9403	0.9437	0.9481	0.9468	0.9418	0.9455
$I_*(0.90)$	0.9357	0.9396	0.9397	0.9397	0.9341	0.9458
$I_{h_1}(0.90)$	0.8918	0.8931	0.9032	0.9038	0.8946	0.9001
$I_{h_2}(0.90)$	0.8906	0.8935	0.9017	0.9020	0.8956	0.8979
$I_{h_3}(0.90)$	0.8904	0.8904	0.8971	0.8956	0.8902	0.8945

2.5 Real Data Analysis

In this section, we revisit the last two data analyses in [16] by constructing the proposed empirical likelihood confidence intervals.

First we apply the proposed empirical likelihood method to predict the conditional $V@R$ of three airlines' stock price changes conditional on the crude oil price changes. This data set contains daily prices of the crude oil, Delta Airline (DAL), American Airline (AMR) and Southwest Airlines (LUV) from May 2007 to July 2008. Let Y_t be the price change of the airline stock (either DAL or AMR or LUV) at day t and X_t be the price change of the crude oil at day t. For DAL and AMR, we fit the following linear model:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t.$$

A different calibration is adopted for LUV to ensure the assumption of uncorrelated errors. Specifically, for LUV we fit the following linear model instead:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + \epsilon_t.$$

Figure 2.1 presents the time series of the residuals and their sample autocorrelations with different numbers of lags, which supports our assumption of uncorrelated errors. It remains open on how to test the independence of errors, which is our future research. Note that the standard Box test is not applicable here with possible infinite variance predicting variables.

In Figures 2.2 and 2.3, we plot the predicted conditional V@R with its 95% intervals with $\alpha = 0.95$ and 0.99 based on the proposed empirical likelihood method. Here the predicted conditional V@R is the maximum empirical likelihood estimate instead of the least squares estimator. From these two figures, we observe that i) the conditional V@R of DAL and AMR share a similar pattern, which suggests that the price changes of both airlines may face a similar risk trend from the crude oil price change; ii) the conditional V@R of

AMR seems to have a larger value and a larger variation than DAL, suggesting that AMR has larger exposures to the risk of crude oil changes; iii) the conditional V@R of LUV shows a more stable pattern over the entire period than that of DAL and AMR, suggesting that the crude oil price change has less influence on the stock return on LUV than that on the other two airlines; iv) the obtained intervals are clearly asymmetric, i.e. the predicted conditional V@Rs do not lie in the exact center of the confidence intervals. This is one well-known advantage of the empirical likelihood inference: it provides data-determined shapes for confidence regions/intervals. In conclusion, these intervals clearly show that uncertainty of the predicted conditional V@R is quite dynamic, and the confidence interval of the conditional V@R is useful in monitoring the changes of risk measures.

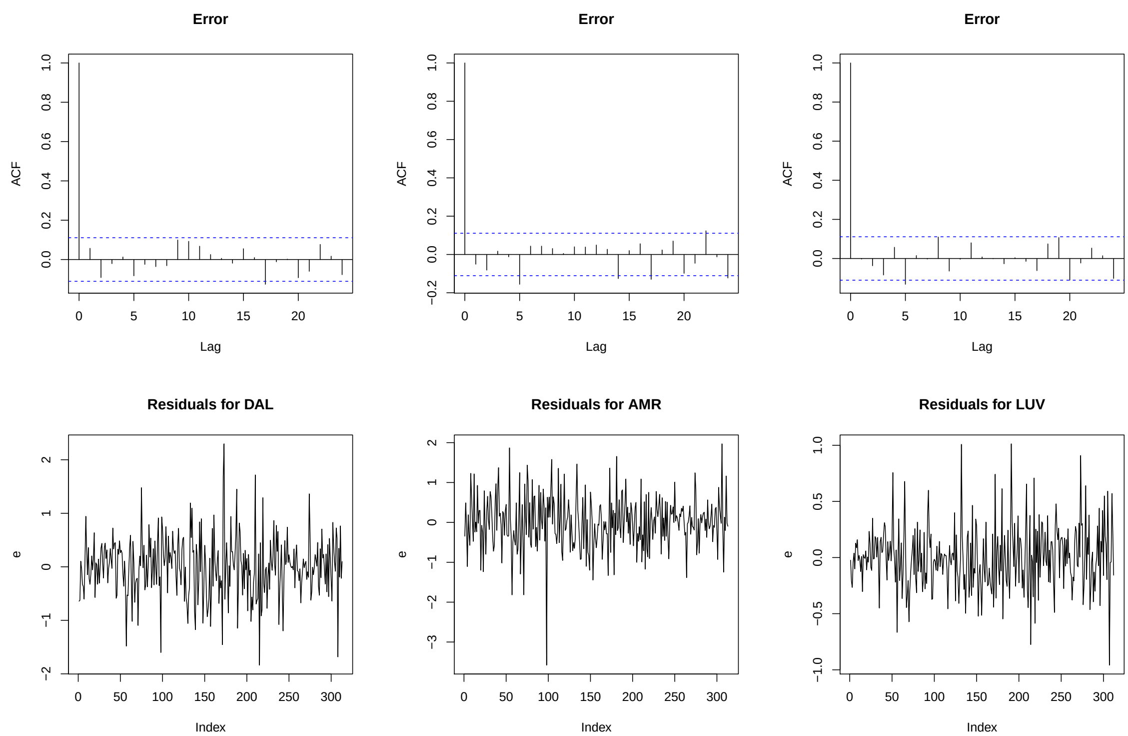


Figure 2.1: Time series of residuals (bottom) and their autocorrelation functions (top) for Delta Airline (left), American Airline (middle) and Southwest Airlines (right).

Second, we apply the proposed method to the International Business Machines Corporation (IBM) data set which contains daily returns on IBM stock and the Standard & Poor's 500 (S&P500) index from December 2005 to December 2015. We calibrate the following

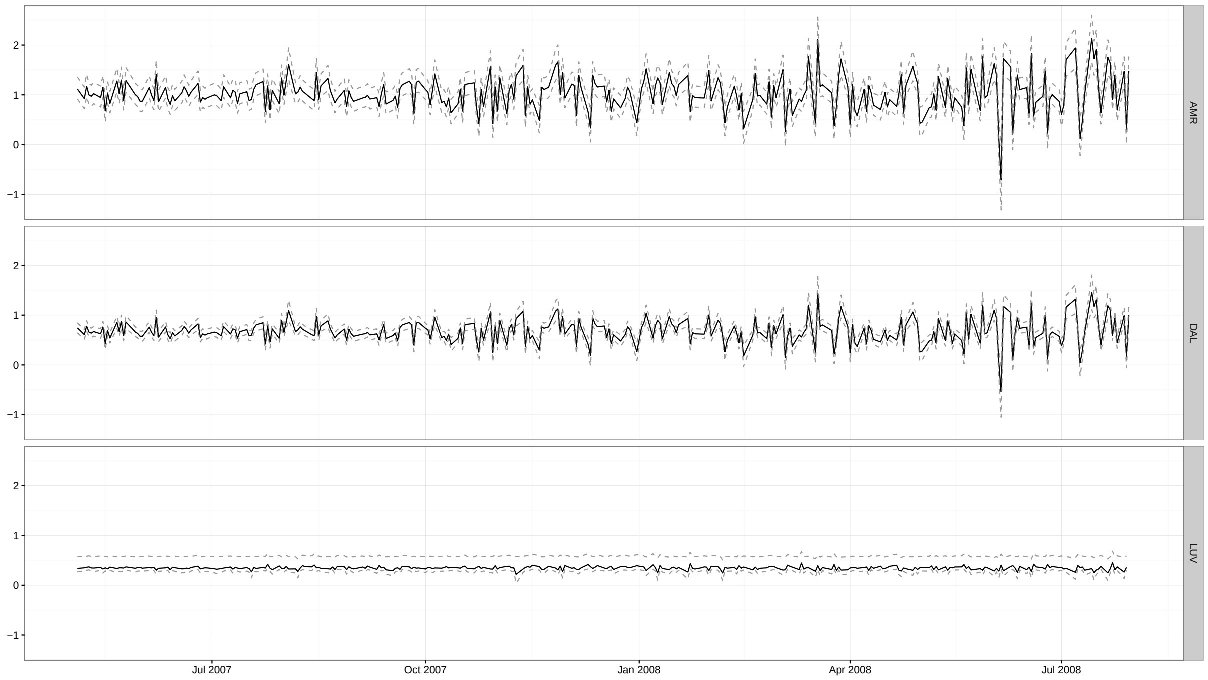


Figure 2.2: Empirical-likelihood asymptotic confidence intervals (dashed) with level 95% and predicted values (solid) for the conditional $V@R$ of the airlines price changes given the crude oil price with $\alpha = 0.95$ and $h = n^{-1/3}$.

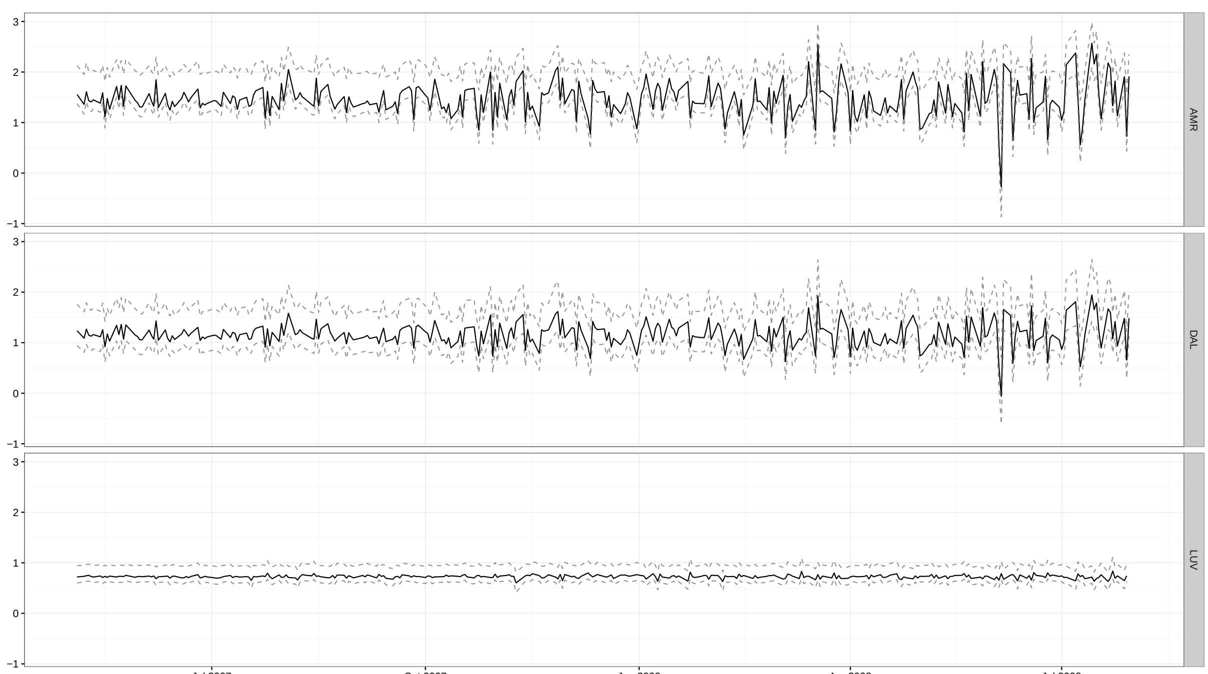


Figure 2.3: Empirical-likelihood asymptotic confidence intervals (dashed) with level 95% and predicted values (solid) for the conditional $V@R$ of the airlines price changes given the crude oil price with $\alpha = 0.99$ and $h = n^{-1/3}$.

linear model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \epsilon_t,$$

where the dependent variable Y_t is the daily log return (in percentage) on IBM, i.e. $Y_t = 100\log(P(t)/P(t-1))$ with $P(t)$ denoting the stock price of IBM at day t, one predicting variable X_t is the log return on S&P500 index and another predicting variable Y_{t-1} is the lagged log return on IBM. We forecast 100 one-day-ahead conditional V@R given the current values of the predicting variables using a rolling window of the previous 1179 days for $\alpha = 0.95$ (which corresponds to predictions at the last 100 days of 2010) and 2437 days for $\alpha = 0.99$ (which corresponds to predictions at the last 100 days of 2015), i.e., $V@R_{\boldsymbol{X}_n}(\alpha)$ with $n = 1179$ for $\alpha = 0.95$ and $n = 2437$ for $\alpha = 0.99$ respectively.

Before constructing the empirical likelihood confidence intervals, we first check the assumption of uncorrelated errors. Figure 2.4 shows the sample autocorrelation function (ACF) of the residuals and the time series of residuals. Although a formal test for uncorrelated errors for a predictive regression is needed, these small values in the ACF plots may indicate the assumption of uncorrelated errors is reasonable. In Figures 2.5 and 2.6, we forecast the one-day-ahead conditional $V@R$ with 95% empirical-likelihood confidence intervals. We observe generally wider confidence intervals for $\alpha = 0.99$ than those for $\alpha = 0.95$, and the time-varying patterns suggest that the proposed interval estimates for the condition V@R are useful in monitoring the changes of risk.

2.6 Proofs

Before proving Theorem 2.2.1, we need the following lemma.

Lemma 2.6.1. *Under conditions A), for a given $M > 0$ there exists $a > 0$ such that*

$$\sup_{|y-\gamma|\le a} \Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{I(\epsilon_t \le y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-F_\epsilon(y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-I(\epsilon_t\le y)+F_\epsilon(y)\}\Big| = o_P(1) \tag{2.5}$$

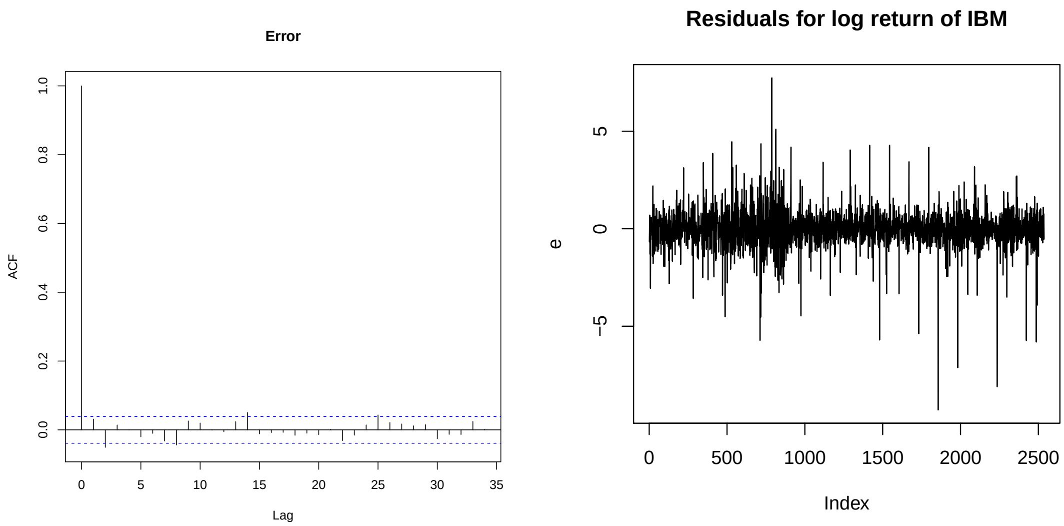


Figure 2.4: Autocorrelation function (left) and the time series (right) of residuals.

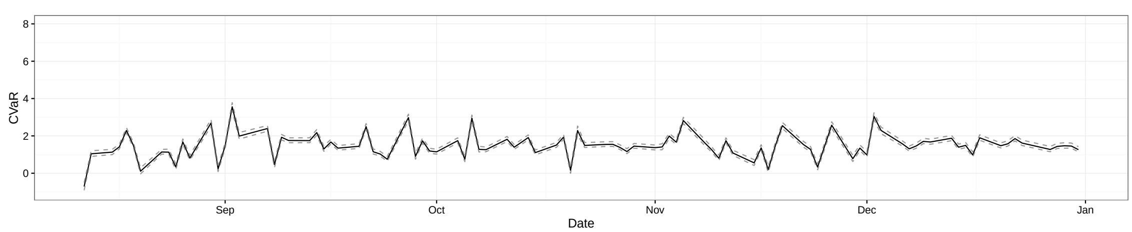


Figure 2.5: Empirical likelihood confidence intervals (dashed) with level 95% and predicted values (solid) for the conditional $V@R$ of daily log return (in percentage) on IBM with $\alpha = 0.95$ and $h = n^{-1/3}$.

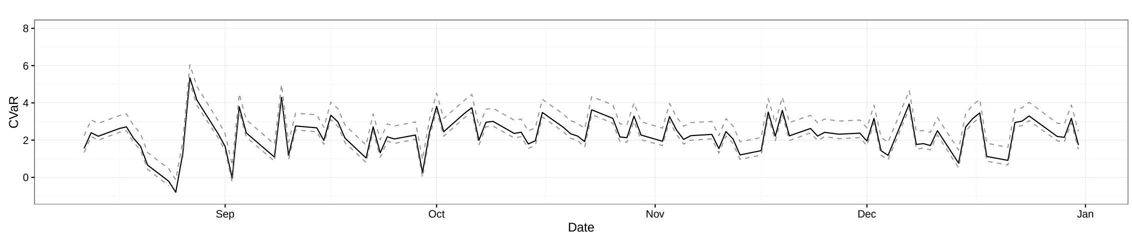


Figure 2.6: Empirical likelihood confidence intervals (dashed) with level 95% and predicted values (solid) for the conditional $V@R$ of log return of IBM stock with $\alpha = 0.99$ and $h = n^{-1/3}$.

and

$$\sup_{|y-\gamma|\le a} |\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{F_\epsilon(y + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y)\}| = O_P(1) \tag{2.6}$$

uniformly for $||\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{k+1})^T|| := \{\sum_{i=1}^{k+1} \delta_i^2\}^{1/2} \le M$, *where* $\gamma = F_\epsilon^{-1}(\alpha)$.

Proof. Note that the condition $E||\boldsymbol{X}_t||^{2+\delta_0} < \infty$ implies that $\max_{1\le t\le n} ||\boldsymbol{X}_t|| = o_P(n^{1/2})$. Since $F_\epsilon'(x)$ is continuous at $x = \gamma$, we can find $a > 0$ such that $\sup_{|x-\gamma|\le 2a} F_\epsilon'(x) < \infty$.

For $d_1 \in (0, \frac{1}{2})$, put $N = [n^{1/2+d_1}]$, $y_i = \gamma - a + \frac{i-1}{N} 2a$ for $i = 1, \cdots, N+1$, and

$$A_t(y) = I(\epsilon_t \le y + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - I(\epsilon_t \le y) + F_\epsilon(y).$$

When $y \in [y_i, y_{i+1}]$, we have

$$\begin{aligned}
& \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y) \\
\le\ & \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y_{i+1}) \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{F_\epsilon(y_{i+1} + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t)\} \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{-I(\epsilon_t \le y_i) + I(\epsilon_t \le y_{i+1}) + F_\epsilon(y_i) - F_\epsilon(y_{i+1})\} \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{-F_\epsilon(y_i) + F_\epsilon(y_{i+1})\}
\end{aligned}$$

and

$$\begin{aligned}
& \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y) \\
\ge\ & \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y_i) \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y_{i+1} + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t)\} \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{I(\epsilon_t \le y_i) - I(\epsilon_t \le y_{i+1}) - F_\epsilon(y_i) + F_\epsilon(y_{i+1})\} \\
& + \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{F_\epsilon(y_i) - F_\epsilon(y_{i+1})\},
\end{aligned}$$

which imply that

$$
\begin{aligned}
& \sup_{|y-\gamma|\le a} |\tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y)| \\
\le\ & \sup_{1\le i\le N+1} |\tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n A_t(y_i)| \\
& + \sup_{1\le i\le N} |\tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{F_\epsilon(y_{i+1} + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t)\}| \\
& + \sup_{1\le i\le N} |\tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{I(\epsilon_t \le y_{i+1}) - I(\epsilon_t \le y_i) - F_\epsilon(y_{i+1}) + F_\epsilon(y_i)\}| \\
& + \sup_{1\le i\le N} \tfrac{1}{\sqrt{n}} \textstyle\sum_{t=1}^n \{F_\epsilon(y_{i+1}) - F_\epsilon(y_i)\} \\
=\ & I_1 + I_2 + I_3 + I_4.
\end{aligned}
\tag{2.7}
$$

Therefore, for any $\Delta > 0$,

$$
\begin{aligned}
& P(I_1 > \Delta, \max_{1\le t\le n} \|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a) \\
\le\ & \textstyle\sum_{i=1}^N P(|\tfrac{1}{\sqrt{n}} \sum_{t=1}^n A_t(y_i)| > \Delta, \max_{1\le t\le n} \|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a) \\
\le\ & \textstyle\sum_{i=1}^N P(|\tfrac{1}{\sqrt{n}} \sum_{t=1}^n A_t(y_i) I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a)| > \Delta) \\
\le\ & \textstyle\sum_{i=1}^N \frac{E(\sum_{t=1}^n A_t(y_i) I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a))^4}{n^2\Delta^4} \\
=\ & \frac{\sum_{i=1}^N \sum_{t=1}^n E\{E(A_t^4(y_i) I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a)|\boldsymbol{Z}_t)\}}{n^2\Delta^4} \\
& + \frac{6\sum_{i=1}^N \sum_{t\ne s} E\{E(A_t^2(y_i) I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a)|\boldsymbol{Z}_t) E(A_s^2(y_i) I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_s\| \le a)|\boldsymbol{Z}_s)\}}{n^2\Delta^4} \\
=\ & O(\frac{\sum_{i=1}^N \sum_{t=1}^n E\{|F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y_i)| I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a)\}}{n^2}) \\
& + O(\frac{\sum_{i=1}^N \sum_{t\ne s} E\{|F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t) - F_\epsilon(y_i)| I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a) |F_\epsilon(y_i + n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_s) - F_\epsilon(y_i)| I(\|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_s\| \le a)\}}{n^2}) \\
=\ & O(Nn^{-3/2}) + O(Nn^{-3} \textstyle\sum_{t\ne s} E(\|\boldsymbol{Z}_t\|) E(\|\boldsymbol{Z}_s\|)) \\
=\ & o(1).
\end{aligned}
\tag{2.8}
$$

Similarly we have

$$
I_3 = o_P(1). \tag{2.9}
$$

It is easy to check that

$$
I_2 I(\max_{1\le t\le n} \|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a) \le \frac{1}{\sqrt{n}} \sum_{t=1}^n \{\sup_{1\le i\le N} (y_{i+1} - y_i)\} \sup_{|x-\gamma|\le 2a} F'_\epsilon(x) = o(1),
$$

i.e., we can show that

$$I_2 = o_P(1) \quad \text{and} \quad I_4 = o_P(1). \tag{2.10}$$

Hence (2.5) follows from (2.7)–(2.10), and (2.6) easily follows from the facts that

$$\sup_{|x-\gamma|\le 2a} F'_\epsilon(x) < \infty \quad \text{and} \quad \max_{1\le t\le n} \|n^{-1/2}\boldsymbol{\delta}^T \boldsymbol{Z}_t\| \le a$$

with probability tending to one. Thus we complete the proof of Lemma 2.6.1. □

Proof of Theorem 2.2.1. Define $F_{n,\widehat{\epsilon}}(x) = \frac{1}{n}\sum_{t=1}^n I(\widehat{\epsilon}_t \le x)$, $F_{n,\epsilon}(x) = \frac{1}{n}\sum_{t=1}^n I(\epsilon_t \le x)$, and write $\theta = V@R_{\boldsymbol{x}}(\alpha)$, $\widehat{\theta} = \widehat{V@R}_{\boldsymbol{x}}(\alpha)$, $\gamma = F_\epsilon^{-1}(\alpha)$ and $\widehat{\gamma} = F_{n,\widehat{\epsilon}}^{-1}(\alpha)$. Then under conditions A), we have

$$\widehat{\theta} = F_{n,\widehat{\epsilon}}^{-1}(\alpha) + \widehat{\boldsymbol{\beta}}^T \boldsymbol{z} \quad \text{and} \quad \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \Omega^{-1}\{\frac{1}{n}\sum_{t=1}^n \epsilon_t \boldsymbol{Z}_t\}\{1 + o_P(1)\} = O_P(n^{-1/2}). \tag{2.11}$$

By writing

$$\begin{aligned}
0 &= F_{n,\widehat{\epsilon}}(\widehat{\gamma}) - \tfrac{[n\alpha]}{n} \\
&= \tfrac{1}{n}\textstyle\sum_{t=1}^n \{I(\epsilon_t \le \widehat{\gamma} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_t) - F_\epsilon(\widehat{\gamma} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_t) - I(\epsilon_t \le \widehat{\gamma}) + F_\epsilon(\widehat{\gamma})\} \\
&\quad + \tfrac{1}{n}\textstyle\sum_{t=1}^n \{I(\epsilon_t \le \widehat{\gamma}) - F_\epsilon(\widehat{\gamma})\} + \tfrac{1}{n}\sum_{t=1}^n \{F_\epsilon(\widehat{\gamma} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_t) - F_\epsilon(\widehat{\gamma})\} \\
&\quad + \{F_\epsilon(\widehat{\gamma}) - F_\epsilon(\gamma)\} + \{\alpha - \tfrac{[n\alpha]}{n}\}, \\
&= II_1 + II_2 + II_3 + II_4 + II_5,
\end{aligned}$$

when $|\widehat{\gamma} - \gamma|$ is small enough, it follows from Lemma 2.6.1 that $II_1 = o_P(n^{-1/2})$ and $II_3 = O_P(n^{-1/2})$. Obviously $II_2 = O_P(n^{-1/2})$ and $II_5 = o(n^{-1/2})$. Therefore $II_4 =$

$O_P(n^{-1/2})$, which implies that $\widehat{\gamma} - \gamma = O_P(n^{-1/2})$, and further

$$\begin{aligned} 0 &= II_2 + II_3 + II_4 + o_P(n^{-1/2}) \\ &= \{F_{n,\epsilon}(\gamma) - F_\epsilon(\gamma)\}\{1 + o_P(1)\} \\ &\quad + F'_\epsilon(\gamma)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\{\tfrac{1}{n}\textstyle\sum_{t=1}^n \boldsymbol{Z}_t\}\{1 + o_P(1)\} \\ &\quad + F'_\epsilon(\gamma)(\widehat{\gamma} - \gamma)\{1 + o_P(1)\} + o_P(n^{-1/2}). \end{aligned}$$

Hence, in conjunction with (2.11)

$$\begin{aligned} \sqrt{n}\{\widehat{\gamma} - \gamma\} &= -\frac{\sqrt{n}\{F_{n,\epsilon}(\gamma) - \alpha\}}{F'_\epsilon(\gamma)} - \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z_1}) + o_P(1) \\ &= -\frac{1}{F'_\epsilon(\gamma)} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n (I(\epsilon_t \le \gamma) - \alpha) - \frac{1}{\sqrt{n}} \sum_{t=1}^n \epsilon_t \boldsymbol{Z}_t^T \Omega^{-1} E(\boldsymbol{Z}_1) + o_P(1) \\ &= -\frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \frac{I(\epsilon_t \le \gamma) - \alpha}{F'_\epsilon(\gamma)} + \epsilon_t \boldsymbol{Z}_t^T \Omega^{-1} E(\boldsymbol{Z}_1) \right\} + o_P(1). \end{aligned}$$

Hence,

$$\begin{aligned} &\sqrt{n}\left(\widehat{V@R}_{\boldsymbol{x}}(\alpha) - V@R_{\boldsymbol{x}}(\alpha)\right) \\ =&\sqrt{n}\{\widehat{\gamma} - \gamma\} + \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T \boldsymbol{z} \\ =&-\frac{\sqrt{n}\{F_{n,\epsilon}(\gamma) - \alpha\}}{F'_\epsilon(\gamma)} + \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T \boldsymbol{z} - \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z_1}) + o_P(1) \\ =&\frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ -\frac{I(\epsilon_t \le \gamma) - \alpha}{F'_\epsilon(\gamma)} + \epsilon_t \boldsymbol{Z}_t^T \Omega^{-1} \boldsymbol{z} - \epsilon_t \boldsymbol{Z}_t^T \Omega^{-1} E(\boldsymbol{Z}_1) \right\} + o_P(1) \\ =:&S_{nn} + o_P(1), \end{aligned}$$

with

$$S_{nt} := \frac{1}{\sqrt{n}} \sum_{m=1}^t \left\{ -\frac{I(\epsilon_m \le \gamma) - \alpha}{F'_\epsilon(\gamma)} + \epsilon_m \boldsymbol{Z}_m^T \Omega^{-1} \boldsymbol{z} - \epsilon_m \boldsymbol{Z}_m^T \Omega^{-1} E(\boldsymbol{Z}_1) \right\}, \quad t = 1, \ldots, n.$$

Define a filtration $\{\mathcal{F}_t\}_{t\ge0} = \{\sigma(\epsilon_1, \ldots, \epsilon_t, \boldsymbol{Z_1}, \ldots, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})\}_{t\ge0}$. Note that $(S_{nt}, \mathcal{F}_t)$ is a zero-mean, squared integrable martingale with differences

$$D_{nt} = \frac{1}{\sqrt{n}}\left\{-\frac{I(\epsilon_t \le \gamma)-\alpha}{F'_\epsilon(\gamma)} + \epsilon_t \boldsymbol{Z}_t^T\Omega^{-1}\boldsymbol{z} - \epsilon_t \boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\} =: D_{nt,1}+D_{nt,2}+D_{nt,3}.$$

Using Markov inequality, we have for all $\varepsilon > 0$

$$\begin{aligned}
&P\left(\max_t |D_{nt}| > \varepsilon\right)\\
\le&\sum_{t=1}^n P\left(|D_{nt,1}| > \varepsilon/2\right) + \sum_{t=1}^n P\left(|D_{nt,2}+D_{nt,3}| > \varepsilon/2\right)\\
\le& n^{-\delta_0/2}\frac{1}{n}\sum_{t=1}^n \frac{(2/F'_\epsilon(\gamma))^{2+\delta_0} + E|\epsilon_t|^{2+\delta_0}E\left\|\boldsymbol{Z}_t\right\|^{2+\delta_0}\left\|\Omega^{-1}\right\|^{2+\delta_0}\left\|\boldsymbol{z}-E(\boldsymbol{Z}_1)\right\|^{2+\delta_0}}{(\varepsilon/2)^{2+\delta_0}}\\
\to& 0, \quad \text{as } n\to\infty.
\end{aligned}$$

In other words, $\max_t |D_{nt}| \xrightarrow{P} 0$. Using the law of large numbers for martingale (e.g. Theorem 2.13 in [42]) with the filtration $\{\mathcal{F}_t\}_{t\ge 0}$, under conditions A) it is easy to verify that, as $n \to \infty$,

$$\begin{aligned}
\sum_{t=1}^n D_{nt}^2 =& \sum_{t=1}^n D_{nt,1}^2 + \sum_{t=1}^n D_{nt,2}^2 + \sum_{t=1}^n D_{nt,3}^2\\
&+ 2\sum_{t=1}^n D_{nt,1}D_{nt,3} + 2\sum_{t=1}^n D_{nt,2}D_{nt,3} + 2\sum_{t=1}^n D_{nt,1}D_{nt,2}\\
\xrightarrow{P}& \omega^2 + \sigma^2\boldsymbol{z}^T\Omega^{-1}\boldsymbol{z} + \sigma^2 E(\boldsymbol{Z}_1)^T\Omega^{-1}E(\boldsymbol{Z}_1) + 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1)\\
&- 2\sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z} - 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z}\\
=&\omega^2 + \sigma^2\boldsymbol{z}^T\Omega^{-1}\boldsymbol{z} + \Delta_1 + \Delta_2.
\end{aligned}$$

Furthermore, it is easy to verify that

$$E\left(\max_t \left\|D_{nt}\right\|^2\right) \le E\left(\sum_{t=1}^n \left\|D_{nt}\right\|^2\right) \le 3\sum_{i=1}^3 E\left(\sum_{t=1}^n \left\|D_{nt,i}\right\|^2\right) = O(1).$$

The theorem then follows from Theorem 3.2 in [42]. $\square$

Before we prove Theorem 2.2.2, we need the following lemma.

Lemma 2.6.2. *Under conditions B), for a given $M > 0$ there exists $a > 0$ such that*

$$\sup_{|y-\gamma|\le a} |\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{I(\epsilon_t \le y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-F_\epsilon(y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-I(\epsilon_t \le y)+F_\epsilon(y)\}| = o_P(1) \tag{2.12}$$

and

$$\sup_{|y-\gamma|\le a} |\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{F_\epsilon(y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-F_\epsilon(y)\}| = O_P(1) \tag{2.13}$$

uniformly for $||\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{k+1})^T|| := \{\sum_{i=1}^{k+1}\delta_i^2\}^{1/2} \le M$, where $\gamma = F_\epsilon^{-1}(\alpha)$.

Proof. Note that

$$\begin{aligned}
& \sup\nolimits_{|y-\gamma|\le a} |\tfrac{1}{\sqrt{n}}\textstyle\sum_{t=1}^{n}\{I(\epsilon_t \le y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-F_\epsilon(y+n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t)-I(\epsilon_t \le y)\\
& +F_\epsilon(y)\}I(|n^{-1/2}\delta_{k+1}X_{k,t}| > a)| \\
\le\ & \tfrac{4}{\sqrt{n}}\textstyle\sum_{t=1}^{n} I(|n^{-1/2}\delta_{k+1}X_{k,t}| > a) \\
=\ & O_P(\sqrt{n}P(|n^{-1/2}\delta_{k+1}X_{k,t}| > a)) \\
=\ & O_P(\sqrt{n}n^{-d/2+(d-1)/4}) = o_P(1).
\end{aligned}$$

Hence the lemma can be proved as that of Lemma 1 by restricting the analysis to the set $||n^{-1/2}\boldsymbol{\delta}^T\boldsymbol{Z}_t|| \le a$. $\square$

Proof of Theorem 2.2.2. Using Lemma 2.6.2, the theorem can be shown in the same way as that of Theorem 2.2.1 by noting that $\widehat{\beta}_k - \beta_k$ has a faster rate of convergence than other parts. We skip details. $\square$

Before proving Theorem 2.3.1, we need some lemmas. For presentation convenience, we shall assume the conditions of Theorem 2.3.1 throughout the below section. Denote $||\cdot||$ as the Frobenius norm (Euclidean norm), that is, for arbitrary matrix (or vector) A

we define $\|A\|$ as the square root of the sum of the squares of its entries. Denote $\xrightarrow{P}$ as convergence in probability and $\xrightarrow{d}$ as convergence in distribution.

Lemma 2.6.3. *As $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{d} N(0, \Sigma_1), \ \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \boldsymbol{W}_t^T(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{P} \Sigma_1,$$

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{P} \Sigma_2.$$

Proof. It is straightforward to verify the above statements by using the weak law of large numbers and the central limit theorem for martingale differences array as in the proof of Theorem 2.2.1 with the same filtration $\{\mathcal{F}_t\}_{t\geq 0} = \{\sigma(\epsilon_1, \ldots, \epsilon_t, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{t+1})\}_{t\geq 0}$; see, e.g., Theorem 2.13 and 3.2 in [42]. Note that for the first part we need to use the fact that

$$E\left(\boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0, \theta^0)|\mathcal{F}_{t-1}\right) = O(h^2) = o(n^{-1/2}), \quad E\left(\boldsymbol{W}_{t,i}(\boldsymbol{\beta}^0, \theta^0)|\mathcal{F}_{t-1}\right) = 0, \quad i = 2, \ldots, k+2,$$

uniformly for $t \geq 1$. □

Lemma 2.6.4. *For all $a \in (0, r_0)$, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-1/a}} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \Sigma_2 \right\| \xrightarrow{P} 0.$$

Proof. By Lemma 2.6.3, it suffices to show

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0, \theta^0)}{\partial \boldsymbol{\beta}} \xrightarrow{P} 0.$$

uniformly in $\{\boldsymbol{\beta}\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq n^{-1/a}\}$. All the statements below hold uniformly for $\boldsymbol{\beta}$ such that $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq n^{-1/a}$; for presentation convenience we do not repeat this argument.

Note that

$$
\begin{aligned}
&\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta},\theta^0)}{\partial \boldsymbol{\beta}}-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0,\theta^0)}{\partial \boldsymbol{\beta}}\right\| \\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{1}{h}g\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t-(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T(\boldsymbol{Z}_t-\boldsymbol{z})}{h}\right)-\frac{1}{h}g\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t}{h}\right)\right\|\frac{\|\boldsymbol{Z}_t-\boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|^2} \\
\leq&\frac{(1+\|\boldsymbol{z}\|)}{n\|\boldsymbol{Z}_t\|}\sum_{t=1}^{n}\left\|\frac{1}{h}g\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t-(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T(\boldsymbol{Z}_t-\boldsymbol{z})}{h}\right)-\frac{1}{h}g\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t}{h}\right)\right\| \\
&\times \mathbb{1}\left[|\epsilon_t-F_\varepsilon^{-1}(\alpha)|\leq h\right] \\
&+\frac{(1+\|\boldsymbol{z}\|)}{n\|\boldsymbol{Z}_t\|}\sum_{t=1}^{n}\left|\frac{1}{h}g\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t-(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T(\boldsymbol{Z}_t-\boldsymbol{z})}{h}\right)\right|\mathbb{1}\left[|\epsilon_t-F_\varepsilon^{-1}(\alpha)|> h\right] \\
=:&T_1+T_2.
\end{aligned}
$$

Applying Taylor expansion to get, for some large $C>0$,

$$
\begin{aligned}
T_1\leq&\sup_x|g'(x)|\cdot\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|\frac{C}{nh^2}\sum_{t=1}^{n}\frac{\|\boldsymbol{Z}_t-\boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|}\mathbb{1}[|\varepsilon_t-F_\varepsilon^{-1}(\alpha)|\leq h] \\
\leq&\frac{C}{n^{1/a}h}\cdot\frac{1}{nh}\sum_{t=1}^{n}\mathbb{1}[|\varepsilon_t-F_\varepsilon^{-1}(\alpha)|\leq h]=o_p(1)\cdot(f(F_\varepsilon^{-1}(\alpha))+o_p(1))=o_p(1),
\end{aligned}
$$

noting that $n^{1/a}h\to\infty$, as $n\to\infty$.

Recall that $g(\cdot)$ is bounded and only supported on $[-1,1]$. For some large $C>0$, we have

$$
\begin{aligned}
T_2 \leq & \frac{C}{\|\boldsymbol{Z}_t\| nh} \sum_{t=1}^{n} \mathbb{1}\left[0 \leq \varepsilon_t - F_\epsilon^{-1}(\alpha) - h \leq \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{Z}_t\| + \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{z}\|\right] \\
& + \frac{C}{\|\boldsymbol{Z}_t\| nh} \sum_{t=1}^{n} \mathbb{1}\left[0 \geq \varepsilon_t - F_\epsilon^{-1}(\alpha) + h \geq -\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{Z}_t\| - \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{z}\|\right] \\
\leq & \frac{C}{\|\boldsymbol{Z}_t\| nh} \sum_{t=1}^{n} \mathbb{1}\left[0 \leq \varepsilon_t - F_\epsilon^{-1}(\alpha) - h \leq n^{-1/a} \|\boldsymbol{Z}_t\| + n^{-1/a} \|\boldsymbol{z}\|\right] \\
& + \frac{C}{\|\boldsymbol{Z}_t\| nh} \sum_{t=1}^{n} \mathbb{1}\left[0 \geq \varepsilon_t - F_\epsilon^{-1}(\alpha) + h \geq -n^{-1/a} \|\boldsymbol{Z}_t\| - n^{-1/a} \|\boldsymbol{z}\|\right] \\
=: & T_{21} + T_{22}.
\end{aligned}
$$

In the sequel we shall only prove that $T_{21} = o_p(1)$; the proof of $T_{22} = o_p(1)$ is completely analogous and therefore omitted. Observe that

$$
\max_{1 \leq t \leq n} n^{-1/a} \|\boldsymbol{Z}_t\| \, \mathbb{1}\left[\|\boldsymbol{Z}_t\| \leq n^{1/r_0}\right] \leq n^{-1/a} \cdot n^{1/r_0} = o(1).
$$

It follows that

$$
\begin{aligned}
T_{21} \leq & \frac{1}{\|\boldsymbol{Z}_t\| nh} \sum_{t=1}^{n} \mathbb{1}\left[0 \leq \varepsilon_t - F_\epsilon^{-1}(\alpha) - h \leq n^{-1/a} \|\boldsymbol{Z}_t\| + n^{-1/a} \|\boldsymbol{z}\|\right] \mathbb{1}\left[\|\boldsymbol{Z}_t\| \leq n^{1/r_0}\right] \\
& + \frac{1}{nh \|\boldsymbol{Z}_t\|} \sum_{t=1}^{n} \mathbb{1}\left[\|\boldsymbol{Z}_t\| > n^{1/r_0}\right] = T_{21,1} + T_{21,2}
\end{aligned}
$$

where, for large $C > 0$ and large n, using Taylor expansion it is easy to show

$$
E(T_{21,1}) = E(E(T_{21,1} | \boldsymbol{Z}_1, \ldots \boldsymbol{Z}_n)) \leq CE\left(\frac{1}{nh} \sum_{t=1}^{n} n^{-1/a} \cdot \frac{\|\boldsymbol{Z}_t\| + \|\boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|}\right) \leq \frac{C(1 + \|\boldsymbol{z}\|)}{n^{1/a} h} = o(1),
$$

and $0 \leq T_{21,2} \leq \frac{1}{n^{1/r_0} h} = o(1)$. Hence, by Markov inequality we have $T_{21} = o_p(1)$. $\square$

Lemma 2.6.5. *Let $a \in (0, 2+\delta_0)$. As $n \to \infty$,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left\| \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0) \right\| = o_P\left(n^{\frac{1}{2+\delta_0}}\right).$$

Proof. Noting that K is a distribution function and $\|\boldsymbol{Z}_t\| \ge 1$,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left| W_{t,1}(\boldsymbol{\beta}, \theta^0) \right| \le 2 = o_P\left(n^{\frac{1}{2+\delta_0}}\right).$$

For $i \ge 2$, noting that $\|\boldsymbol{Z}_t\| \ge 1$ and $\|\boldsymbol{Z}_t\| \ge |X_{t,i}|$,

$$\begin{aligned}
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left| W_{t,i}(\boldsymbol{\beta}, \theta^0) \right| &= \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left| \epsilon_t \frac{X_{t,i-2}}{\|\boldsymbol{Z}_t\|^2} - \left(\boldsymbol{\beta} - \boldsymbol{\beta}^0\right) \frac{\boldsymbol{Z}_t X_{t,i-2}}{\|\boldsymbol{Z}_t\|^2} \right| \\
&\le \max_{1 \le t \le n} |\epsilon_t| + n^{-\frac{1}{a}} = o_P\left(n^{\frac{1}{2+\delta_0}}\right),
\end{aligned}$$

where in the last step we use $\max_{1 \le t \le n} |\epsilon_t| = o_P\left(n^{\frac{1}{2+\delta_0}}\right)$ by Lemma 3 in [43]. Hence,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left\| \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0) \right\| \le \sum_{i=1}^{k+2} \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \max_{1 \le t \le n} \left| W_{t,i}(\boldsymbol{\beta}, \theta^0) \right| = o_P\left(n^{\frac{1}{2+\delta_0}}\right).$$

□

Lemma 2.6.6. *For all $a > 0$, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-\frac{1}{a}}} \left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0) \boldsymbol{W}_t^T(\boldsymbol{\beta}, \theta^0) - \Sigma_1 \right\| \xrightarrow{P} 0.$$

Proof. Recall from Lemma 2.6.3, as $n \to \infty$,

$$\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \boldsymbol{W}_t^T(\boldsymbol{\beta}^0, \theta^0) - \Sigma_1 \xrightarrow{P} 0.$$

It suffices to verify that

$$\sup_{\|\beta-\beta^0\|\le n^{-\frac{1}{a}}} \frac{1}{n}\sum_{t=1}^{n}\left(W_{t,i}(\boldsymbol{\beta},\theta^0)-W_{t,i}(\boldsymbol{\beta}^0,\theta^0)\right)^2 \xrightarrow{P} 0, \quad i=1,\ldots,k+2.$$

This is trivial for $i\ge 2$ since uniformly for $\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\le n^{-\frac{1}{a}}$

$$\frac{1}{n}\sum_{t=1}^{n}\left(W_{t,i}(\boldsymbol{\beta},\theta^0)-W_{t,i}(\boldsymbol{\beta}^0,\theta^0)\right)^2=\frac{1}{n}\sum_{t=1}^{n}\left((\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T\boldsymbol{Z}_t\frac{X_{t,i}}{\|\boldsymbol{Z}_t\|^2}\right)^2\le\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|^2\xrightarrow{P}0.$$

On the other hand, taking some $r>a$ and noting that $\|\boldsymbol{Z}_t\|\ge 1$,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(W_{t,1}(\boldsymbol{\beta},\theta^0)-W_{t,1}(\boldsymbol{\beta}^0,\theta^0)\right)^2\\
\le&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t-(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T(\boldsymbol{Z}_t-\boldsymbol{z})}{h}\right)-K\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t}{h}\right)\right\}^2\\
&\frac{1}{\|\boldsymbol{Z}_t\|^4}I(\|\boldsymbol{Z}_t\|\le n^{\frac{1}{r}})+\frac{4}{n}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|^4}I(\|\boldsymbol{Z}_t\|>n^{\frac{1}{r}})\\
\le&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t+n^{\frac{1}{r}-\frac{1}{a}}+n^{-\frac{1}{a}}\|\boldsymbol{z}\|}{h}\right)\right.\\
&\left.-K\left(\frac{F_\epsilon^{-1}(\alpha)-\epsilon_t-n^{\frac{1}{r}-\frac{1}{a}}-n^{-\frac{1}{a}}\|\boldsymbol{z}\|}{h}\right)\right\}^2+4n^{-\frac{2}{r}}\\
\le&\frac{4}{n}\sum_{t=1}^{n}I\left[-h-n^{\frac{1}{r}-\frac{1}{a}}-n^{-\frac{1}{a}}\|\boldsymbol{z}\|<\epsilon_t-F_\epsilon^{-1}(\alpha)<h+n^{\frac{1}{r}-\frac{1}{a}}+n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right]+4n^{-\frac{2}{r}}\\
=&\left|F_\epsilon\left(F_\epsilon^{-1}(\alpha)+h+n^{\frac{1}{r}-\frac{1}{a}}+n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right)-F_\epsilon\left(F_\epsilon^{-1}(\alpha)-h-n^{\frac{1}{r}-\frac{1}{a}}-n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right)\right|+o_p(1)\\
=&o_p(1)
\end{aligned}$$

uniformly in $\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\le n^{-\frac{1}{a}}\}$. □

The following lemma establishes the quadratic expansion of the empirical likelihood ratio function $-2\log L(\cdot,\theta^0)$ around the true parameters $\boldsymbol{\beta}^0$.

Lemma 2.6.7. *Let* $a \in (2, \min\{2+\delta_0/(3+\delta_0), r_0\})$. *For all* $\boldsymbol{\beta}$ *such that* $\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq n^{-1/a}$,

$$-2\log L(\boldsymbol{\beta}, \theta^0) = \boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \Sigma_2 \boldsymbol{\nu} + 2\boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + \mathbb{W}_n \Sigma_1^{-1} \mathbb{W}_n + o_p(1) + o_p(\|\boldsymbol{\nu}\|) + o_p(\|\boldsymbol{\nu}\|^2) \tag{2.14}$$

with $\mathbb{W}_n := \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{W}_t(\beta^0, \theta^0)$ *and* $\boldsymbol{\nu} := \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)$.

Proof. All the statements below hold uniformly in $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq n^{-1/a}\}$; for presentation convenience we do not repeat this argument.

With Lemmas 2.6.4 and 2.6.6, similar to the proof of (2.14) in [43] (see also (A.1) in [2]), we can show that

$$\boldsymbol{\lambda}(\boldsymbol{\beta}) = O_p(n^{-1/a}), \tag{2.15}$$

and, in conjunction with Lemma 2.6.5, we may write

$$\begin{aligned}
\boldsymbol{\lambda}(\boldsymbol{\beta}) =& \Sigma_1^{-1} \left(\frac{1}{n} \sum_{t=1}^n \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) + \Sigma_2(\boldsymbol{\beta} - \boldsymbol{\beta}^0) \right) + o_p(n^{-1/2}) + o_p(\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\|) \\
=& n^{-1/2} \Sigma_1^{-1} (\mathbb{W}_n + \Sigma_2 \boldsymbol{\nu}) + o_p(n^{-1/2}) + o_p(n^{-1/2}\|\boldsymbol{\nu}\|)
\end{aligned} \tag{2.16}$$

as in the proof of (2.17) in [43] by noting that

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n \left\|\boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\right\|^3 \|\boldsymbol{\lambda}(\boldsymbol{\beta})\| \frac{1}{|1 - \boldsymbol{\lambda}(\boldsymbol{\beta})^T \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)|} \\
\leq & \max_{1 \leq t \leq n} \left\|\boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\right\| \cdot \frac{1}{n} \sum_{t=1}^n \left\|\boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\right\|^2 \|\boldsymbol{\lambda}(\boldsymbol{\beta})\|^2 \cdot \max_{1 \leq t \leq n} \frac{1}{|1 - \boldsymbol{\lambda}(\boldsymbol{\beta})^T \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)|} \\
=& o_p(n^{1/(2+\delta_0)}) \cdot O_p(1) \cdot O_p(n^{-2/a}) \cdot O_p(1) = o_p(n^{-1/2}).
\end{aligned}$$

Now we may expand

$$
\begin{aligned}
&-2\log L(\boldsymbol{\beta},\theta)\\
=&2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)-n\cdot\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right)^2+\frac{2n}{3}\frac{1}{n}\sum_{t=1}^{n}\frac{(\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0))^3}{(1+\delta_t)}\\
=:&S_1-S_2+S_3
\end{aligned}
$$

where, recalling (2.15) and Lemma 2.6.5,

$$
\max_{1\le t\le n}|\delta_t|\le\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|\max_{1\le t\le n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|=O_p(n^{-1/a})\cdot o_p(n^{1/(2+\delta_0)})=o_p(1)
$$

and therefore, in conjunction with Lemma 2.6.6, S_3 has a norm bounded by

$$
\begin{aligned}
&n\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|^3\cdot\max_{1\le t\le n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|^2\max_{1\le t\le n}\frac{1}{|1+\delta_t|}\\
=&n\cdot O_p(n^{-3/a})\cdot o_p(n^{\frac{1}{2+\delta_0}})\cdot O_p(1)\cdot O_p(1)=o_p(1).
\end{aligned}
$$

Write

$$
S_1=2n^{1/2}\boldsymbol{\lambda}(\boldsymbol{\beta})^T\mathbb{W}_n+2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T\left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)-\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta}^0,\theta^0)\right\}=:S_{11}+S_{12}.
$$

Substituting (2.16) in above equation yields that

$$
S_{11}=2\mathbb{W}_n^T\Sigma_1^{-1}\mathbb{W}_n+2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n+o_p(1)+o_p(\|\boldsymbol{\nu}\|)
$$

and, together with Lemma 2.6.4 and Taylor expansion,

$$
\begin{aligned}
S_{12}=&2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T(\Sigma_2+o_p(1))(\boldsymbol{\beta}-\boldsymbol{\beta}^0)\\
=&2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n+2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\Sigma_2\boldsymbol{\nu}+o_p(\|\boldsymbol{\nu}\|)+o_p(\|\boldsymbol{\nu}\|^2).
\end{aligned}
$$

Similarly, substituting (2.16) and using Lemma 2.6.6 yields that

$$
\begin{aligned}
S_2 =& \boldsymbol{\lambda}(\boldsymbol{\beta})^T \left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\boldsymbol{W}_t^T(\boldsymbol{\beta}, \theta^0)\right] \boldsymbol{\lambda}(\boldsymbol{\beta}) \\
=& \boldsymbol{\lambda}(\boldsymbol{\beta})^T(\Sigma_1 + o_p(1))\boldsymbol{\lambda}(\boldsymbol{\beta}) \\
=& \mathbb{W}_n^T\Sigma_1^{-1}\mathbb{W}_n + 2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n + \boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\Sigma_2\boldsymbol{\nu} + o_p(1) + o_p(\|\boldsymbol{\nu}\|) + o_p(\|\boldsymbol{\nu}\|^2).
\end{aligned}
$$

The lemma follows by collecting the substitutions above. □

Lemma 2.6.8. *Let* $a \in (2, \min\{2 + \delta_0/(3 + \delta_0), r_0\})$. *With probability tending to one,* $L(\boldsymbol{\beta}, \theta^0)$ *attains its maximum at some point* $\widetilde{\boldsymbol{\beta}}$ *in the interior of the ball* $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \le n^{-1/a}$ *and, furthermore,*

$$
\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = -(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2\Sigma_1^{-1}\mathbb{W}_n + o_p(1). \tag{2.17}
$$

where $\mathbb{W}_n$ *is defined in Lemma 2.6.7.*

Proof. The existence of $\widetilde{\boldsymbol{\beta}}$ can be proved similar as Lemma 1 in [2] using Lemmas 2.6.3, 2.6.4 and 2.6.6 here, and therefore omitted. The proof of the second part is a (slight) modification of that of Theorem 2 in [44]. We define

$$
\Gamma_n(\boldsymbol{\nu}, \theta^0) = \frac{1}{n}\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\Sigma_2\boldsymbol{\nu} + \frac{2}{n}\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n + o_p(n^{-1}). \tag{2.18}
$$

and take the minimum point of the first two terms $\boldsymbol{\nu}^* = -(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n$. Suppose $\widetilde{\boldsymbol{\nu}} = \boldsymbol{\nu}^*$ only holds for finite n; otherwise, the theorem is proved. Since $\boldsymbol{\beta}^0$ is the interior of the ball $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \le n^{-1/a}$, for large n we have

$$
\Gamma_n(\boldsymbol{\nu}^*, \theta^0) \le \Gamma_n(\widetilde{\boldsymbol{\nu}}, \theta^0).
$$

Applying (2.18) twice in the last expression, consolidating terms, and using the facts that Σ_1 is positive definite (therefore $\Sigma_2^T\Sigma_1^{-1}\Sigma_2$ is positive definite) and $\mathbb{W}_n = O_p(1)$ by Lemma

2.6.3, to get

$$
\begin{aligned}
0 &\leq \left(\widetilde{\boldsymbol{\nu}} + \left(\Sigma_2^T\Sigma_1^{-1}\Sigma_2\right)^{-1}\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n\right)^T \Sigma_2^T\Sigma_1^{-1}\Sigma_2 \left(\widetilde{\boldsymbol{\nu}} + \left(\Sigma_2^T\Sigma_1^{-1}\Sigma_2\right)^{-1}\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n\right) \\
&\leq o_p(1) + o_p(\|\widetilde{\boldsymbol{\nu}}\|) + o_p(\|\widetilde{\boldsymbol{\nu}}\|^2) = o_p(\|\widetilde{\boldsymbol{\nu}} + \left(\Sigma_2^T\Sigma_1^{-1}\Sigma_2\right)^{-1}\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n\|^2),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\nu}} := \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$. Hence, $\widetilde{\boldsymbol{\nu}} + \left(\Sigma_2^T\Sigma_1^{-1}\Sigma_2\right)^{-1}\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n = o_p(1)$ and the lemma follows. $\square$

Proof of Theorem 2.3.1. Recall the maximum empirical likelihood estimator $\widetilde{\boldsymbol{\beta}}$ from Lemma 2.6.8. Substituting (2.17) into (2.14) in Lemma 2.6.7 yields that

$$
-2\log L^P(\theta^0) = -2\log L(\widetilde{\boldsymbol{\beta}}, \theta^0) = \mathbb{W}_n\Sigma_1^{-1/2}D\Sigma_1^{-1/2}\mathbb{W}_n + o_p(1)
$$

where

$$
D := \left\{I_{k+2} - \Sigma_1^{-1/2}\Sigma_2(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}\right\}
$$

and I_{k+2} denotes the identity $(k+2)\times(k+2)$ matrix. Hence the theorem follows from the facts that D is symmetric, idempotent, and

$$
\begin{aligned}
& tr(I_{k+2} - \Sigma_1^{-1/2}\Sigma_2(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}) \\
= \;& k+2 - tr((\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_2) \\
= \;& k+2 - tr((\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1}\Sigma_2) \\
= \;& k+2-(k+1) = 1
\end{aligned}
$$

and $\Sigma_1^{-1/2}\mathbb{W}_n \xrightarrow{d} N(0, I_{k+2})$ by Lemma 2.6.3. $\square$

CHAPTER 3

INFERENCE FOR RELATIVE RISK MEASURE

This Chapter proposes a relative risk measure, which is sensitive to the market comovement, for monitoring systemic risk from regulators' point of view. The asymptotic normality of a nonparametric estimator and its smoothed version are established when the observations are independent. In order to effectively construct an interval without complicated asymptotic variance estimation, a jackknife empirical likelihood inference procedure based on the smoothed nonparametric estimation is provided with a Wilks type of result in case of independent observations. When data follow from AR-GARCH models, the relative risk measure with respect to the errors becomes useful and so we propose a corresponding nonparametric estimator. A simulation study and real-life data analysis show that the proposed relative risk measure is useful in monitoring systemic risk. The content of this chapter is based on Y. He, Y. Hou, L. Peng and J. Sheng (2016). Statistical Inference for a Relative Risk Measure. *Preprint*.

3.1 Introduction

The recent financial crisis has highlighted the impact of systemic risk on the stability of the financial system. On the quantitative side, academics and policy makers have called for advanced statistical tools to measure systemic risk. Although a formal definition of systemic risk does not exist arguably, it is commonly agreed that systemic risk involves the co-movement of several key financial variables. Many measures have been proposed in the literature on banking industry; see the excellent review paper [45]. Studies on systemic risk in the insurance/reinsurance industry have started to attract attention too. For example, [46] used public data to investigate the Granger causality effect between banks and insurers by using some existing systemic risk measures, [47] used network to analyze global insurers, and [48] examined the relationship between insurance activities and systemic risk. The existing literature provides a diversified and controversial picture of the systemic relevance of the insurance/reinsurance industry; see [49, 46, 50, 51].

From regulators' point of view, having risk measures from each agency does not help understand/measure systemic risk at all. Instead, it would be more meaningful to have some relative risk measures reported by each agency with respect to a common benchmark, and hence, regulators could focus on further modeling, analyzing and monitoring those agencies with a larger relative risk. Therefore, an interesting question becomes i) how to define a relative risk measure, which should be quite sensitive to the market co-movement for the purpose of studying systemic risk, and ii) how to infer such a relative risk measure.

Let X and Y denote the random losses, respectively, on an individual portfolio and some benchmark variable, say, a financial market index with joint continuous distribution function $F(x, y)$. Consider the commonly employed expected shortfall risk measure, at level $\alpha \in (0, 1)$, defined as

$$ES_\alpha(X) = E[X|F_1(X) > 1 - \alpha] \quad \text{and} \quad ES_\alpha(Y) = E[Y|F_2(Y) > 1 - \alpha],$$

where F_1 and F_2 are the marginal distributions of X and Y given by $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. A quick way to compare these two risk measures is to look at their ratio $ES_\alpha(X)/ES_\alpha(Y)$ (or difference). However this ratio or difference is invariant to the copula of X and Y, i.e., it is irrelevant to the market comovement. To capture the extreme dependence between X and Y, recently, [52] proposed to multiply the above ratio by the coefficient of (upper) tail dependence

$$\lambda = \lim_{t \downarrow 0} P(F_1(X) > 1 - t|F_2(Y) > 1 - t),$$

which is proposed by [53] and [54] and widely studied by many others in modeling extreme events. Since the coefficient of tail dependence is defined in a limiting way, nonparametric estimator for it has a slower rate of convergence than that for the ratio of expected shortfalls. This means the variability of nonparametrically estimating the ratio of expected shortfalls does not impact the asymptotic behavior of the nonparametric estimator for the defined relative risk in [52]. In other words, nonparametric estimator for the proposed relative risk measure in [52] is not sensitive to the variability of individual risks, which is not good to serve as a systemic risk. Indeed in the empirical study, [52] computes both tail sensitivity and risk measures at the same level. That is, [52] defined the following *relative risk* measure

$$\rho_\alpha = \rho_\alpha(X, Y) = P(F_1(X) > 1 - \alpha | F_2(Y) > 1 - \alpha) \frac{ES_\alpha(X)}{ES_\alpha(Y)}.$$

Although [52] mentioned a nonparametric estimator for the above relative risk, which is called tail risk by them, there is no any theoretical justification. In general when people talk about tail risk, it usually means the level $\alpha = \alpha(n)$ goes to zero as the sample size $n \to \infty$. Also it is important to quantify the uncertainty of a risk measure in risk management. Hence this chapter aims to derive asymptotic limit for a nonparametric estimator and its smoothed version of the above relative risk measure and to provide an effective way to construct an interval by considering either a fixed level or an intermediate level.

In order to implement the above relative risk measure ρ_α at a fixed level $\alpha \in (0, 1)$, this chapter first proposes a nonparametric estimator and its smoothing version and derives an asymptotic normality result based on independent observations. Since the asymptotic variance is quite complicated, we further investigate the possibility of employing an empirical likelihood method to construct a confidence interval since the empirical likelihood method has shown to be quite useful in interval estimation and hypothesis testing. We refer to [1] for an overview of the method. Quantifying uncertainty is important in risk management, and applications of empirical likelihood methods to risk measures have appeared in [39, 55, 56, 57]. In general, an empirical likelihood method is quite effective for linear functionals and requires linearization for a nonlinear functional by introducing some nuisance parameters. Since it is hard to linearize the proposed relative risk measure, we propose to employ the smoothed jackknife empirical likelihood method to construct a confidence interval for the proposed relative risk measure as motivated by the study for copulas and tail copulas in [58] and [59]. Note that smoothed jackknife empirical likelihood method is a generalization of the jackknife empirical likelihood method proposed by [60] for dealing with nonlinear functionals, and smoothing is generally necessary for a non-smoothing nonlinear functional.

When the level α is close to zero, which is a key interest of regulators, and the sample size n is not large enough, it is useful to model α as a function of n. This is generally classified as two situations: intermediate level (i.e., $\alpha = \alpha_n \to 0$ and $\alpha_n n \to \infty$ as $n \to \infty$) and extreme level (i.e., $\alpha = \alpha_n \to 0$ and $n\alpha \to c \in [0, \infty)$ as $n \to \infty$). Such a divergent level relates to the so-called tail risk in financial econometrics, which plays an important role in risk management; see, e.g., [61]. In general, an extreme level requires extrapolating outside the data range. Here we focus on the intermediate level and extend the above study for a fixed level to this case too. Like quantile estimation, we show that nonparametric estimation for the proposed relative risk has a different asymptotic limit for a fixed level and an intermediate level. However, the proposed smoothed jackknife empirical likelihood method gives a unified interval for ρ_α regardless of the level being fixed or intermediate.

The above study is based on independent observations. When data follow from time series models, it becomes more meaningful to consider the relative risk measure of errors. Motivated by our real data applications, we consider an AR-GARCH model and propose a profile empirical likelihood method to construct an interval for the relative risk measure of errors without estimating the complicated asymptotic variance of a nonparametric estimator.

We organize this chapter as follows. Section 3.2 presents our nonparametric estimation procedure, jackknife empirical likelihood method, and asymptotic results based on independent data. When data follow from AR-GARCH models, a profile empirical likelihood method is proposed to construct an interval for the relative risk measure of errors in Section 3.3. A simulation study is carried out in Section 3.4, and a data analysis in finance is provided in Section 3.5 to demonstrate the usefulness of the proposed relative measure in monitoring systemic risk. All proofs are deferred to Section 3.6.

3.2 Main Results For Independent Data

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed random vectors with distribution function $F(x, y)$ and marginals $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. Order the X_i's as $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ and the Y_i's as $Y_{1:n} \le Y_{2:n} \le \ldots \le Y_{n:n}$. Define the survival functions $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$ and quantile functions $Q_i(\cdot) = F_i^{\leftarrow}(\cdot)$ for $i = 1, 2$, where $F_i^{\leftarrow}$ denotes the (generalized) inverse function of F_i. The empirical survival functions are given by

$$\bar{F}_{n1}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > x\right), \quad \bar{F}_{n2}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(Y_i > y\right), \quad x, y \in \mathbb{R}.$$

We introduce the so-called *survival* copula function

$$C(u, v) = P(\bar{F}_1(X) < u, \bar{F}_2(Y) < v), \quad u, v \in [0, 1],$$

and we can rewrite

$$\rho_\alpha = \frac{1}{\alpha} C(\alpha, \alpha) \frac{ES_\alpha(X)}{ES_\alpha(Y)}.$$

Substituting the right-hand-side components by their empirical counterparts yields our nonparametric estimator

$$\widetilde{\rho}_\alpha = \widetilde{\rho}_\alpha(X, Y) = \frac{1}{\alpha} \widetilde{C}(\alpha, \alpha) \frac{\widetilde{ES}_\alpha(X)}{\widetilde{ES}_\alpha(Y)},$$

where, with $\lceil \cdot \rceil$ denoting the ceiling function,

$$\widetilde{C}(\alpha, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left[\bar{F}_{n1}(X_i) < \alpha, \bar{F}_{n2}(Y_i) < \alpha\right],$$

$$\widetilde{ES}_\alpha(X) = \frac{1}{n\alpha} \sum_{i=1}^{n} X_i \mathbb{1}\left[X_i > X_{n-\lceil n\alpha \rceil:n}\right], \quad \widetilde{ES}_\alpha(Y) = \frac{1}{n\alpha} \sum_{i=1}^{n} Y_i \mathbb{1}\left[Y_i > Y_{n-\lceil n\alpha \rceil:n}\right].$$

Like smooth distribution (copula) estimation, we may consider a smooth version of the above nonparametric estimation. More specifically, with some density function k, its distribution function $K(x) = \int_{-\infty}^{x} k(s)ds$ and bandwidth $h = h(n) > 0$, a smoothed estimator of ρ_α is given by

$$\widehat{\rho}_\alpha = \frac{1}{\alpha} \widehat{C}(\alpha, \alpha) \frac{\widehat{ES}_\alpha(X)}{\widehat{ES}_\alpha(Y)},$$

where

$$\begin{cases} \widehat{C}(\alpha, \alpha) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right), \\ \widehat{ES}_\alpha(X) = \frac{1}{n\alpha} \sum_{i=1}^{n} (X_i - X_{n-\lceil n\alpha \rceil:n}) K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) + X_{n-\lceil n\alpha \rceil:n}, \\ \widehat{ES}_\alpha(Y) = \frac{1}{n\alpha} \sum_{i=1}^{n} (Y_i - Y_{n-\lceil n\alpha \rceil:n}) K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right) + Y_{n-\lceil n\alpha \rceil:n}. \end{cases}$$

To establish the asymptotic normality of $\widetilde{\rho}_\alpha$ and $\widehat{\rho}_\alpha$ for a fixed level $\alpha \in (0, 1)$, we will need the following regularity conditions.

Assumption 3.2.1 (Fixed level). [(1.a)]

1. For $j = 1, 2$, Q_j is Lipschitz continuous in a neighborhood of $1 - \alpha$ with $Q_j(1 - \alpha) > 0$, and F_j is strictly increasing and differentiable in a neighborhood of $Q_j(1 - \alpha)$. Moreover, for some $\delta > 0$, $\mathbb{E}(X_+^{2+\delta}) < \infty$ and $\mathbb{E}(Y_+^{2+\delta}) < \infty$, where $x_+ = \max\{x, 0\}$.

2. C has continuous first-order derivatives $C_1(x, \alpha) = \frac{\partial C(x,\alpha)}{\partial x}$ and $C_2(\alpha, y) = \frac{\partial C(\alpha,y)}{\partial y}$ in a neighborhood of, respectively, $x = \alpha$ and of $y = \alpha$.

Assumption (1.a) contains standard conditions, which require underlying local continuity of the marginal distributions together with finite moments for the positive losses; see, e.g., [14]. Assumption (1.b) ensures the application of the standard empirical copula process result; see, e.g., Section V in [62]. Below is an asymptotic normality result, where '$\xrightarrow{d}$' denotes convergence in distribution and '$\xrightarrow{P}$' denotes convergence in probability.

Theorem 3.2.1 (Fixed level)**.** *For an $\alpha \in (0,1)$ satisfying $C(\alpha,\alpha) > 0$, Assumption 3.2.1 implies that*

$$\sqrt{n\alpha}\left(\frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1\right) \xrightarrow{d} N\left(0, \sigma_\alpha^2\right),$$

as $n \to \infty$, with $\sigma_\alpha^2 = Var\left(\Lambda_\alpha + \Theta_{\alpha,1} - \Theta_{\alpha,2}\right)$ and the zero-mean Gaussian random variables

$$\Lambda_\alpha = \frac{\sqrt{\alpha}}{C(\alpha,\alpha)}\left\{B_C(\alpha,\alpha) - C_1(\alpha,\alpha)B_C(\alpha,1) - C_2(\alpha,\alpha)B_C(1,\alpha)\right\},$$
$$\Theta_{\alpha,1} = -\frac{\frac{1}{\sqrt{\alpha}}\int_0^1 B_C(\alpha x,1)dQ_1(1-\alpha x)}{ES_\alpha(X)}, \quad \Theta_{\alpha,2} = -\frac{\frac{1}{\sqrt{\alpha}}\int_0^1 B_C(1,\alpha y)dQ_2(1-\alpha y)}{ES_\alpha(Y)}.$$

Here, B_C is a C-Brownian bridge, i.e., a zero-mean Gaussian process with covariance function

$$\mathbb{E}(B_C(u_1,v_1)B_C(u_2,v_2)) = C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1,v_1)C(u_2,v_2), \quad (u_1,v_1),(u_2,v_2) \in [0,1]^2.$$

Furthermore, if k is a symmetric density with support [-1,1] and bounded first derivative and the bandwidth $h = h(n) > 0$ satisfies

$$nh^2 \to \infty \quad \textit{and} \quad nh^4 \to 0,$$

then we have that, as $n \to \infty$,

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha - \widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{P} 0.$$

Theorem 3.2.1 states that, under weak regularity conditions, both the non-smoothed estimator $\widetilde{\rho}_\alpha$ and smoothed estimator $\widehat{\rho}_\alpha$ are asymptotically normal with the same limiting distribution.

When α is close to zero (but not extremely), as discussed in Section 1, it is often useful to model α as an intermediate sequence of n in such a way that $\alpha = \alpha_n \to 0$ and $n\alpha_n \to \infty$, as $n \to \infty$. For the study of an intermediate level α, in the context of extreme value theory, one needs some conditions on the tail behavior of the underlying variables as follows.

Assumption 3.2.2 (Intermediate level). [(2.a)]

1. For some $\gamma_j \in (0, 1/2)$, $\beta_j \le 0$ and function A_j with a constant sign near infinity,

$$\lim_{t\to\infty} \frac{1}{A_j(1/\bar{F}_j(t))} \left(\frac{\bar{F}_j(tx)}{\bar{F}_j(t)} - x^{-1/\gamma_j} \right) = x^{-1/\gamma_j} \frac{x^{\beta_j/\gamma_j} - 1}{\gamma_j \beta_j}, \quad x > 0, \tag{3.1}$$

   for all $j = 1, 2$.

2. There exists a function $R : (0, \infty)^2 \to [0, \infty)$ such that

$$\lim_{t\to\infty} tC(t^{-1}x, t^{-1}y) = R(x, y), \quad (x, y) \in (0, \infty)^2, \tag{3.2}$$

   and it has continuous first-order derivatives $R_1(x, y) = \frac{\partial R(x,y)}{\partial x}$ and $R_2(x, y) = \frac{\partial R(x,y)}{\partial y}$ on a neighborhood of $(1, 1)$.

3. The function C has first-order derivatives $C_1(x, y) = \frac{\partial C(x,y)}{\partial x}$ and $C_2(x, y) = \frac{\partial C(x,y)}{\partial y}$ on $(0, \delta)^2$ for some $\delta > 0$, and, as $t \to \infty$,

$$\sup_{x,y\in(1-\delta,1+\delta)} \left| C_1(t^{-1}x, t^{-1}y) - R_1(x, y) \right| \to 0,$$
$$\sup_{x,y\in(1-\delta,1+\delta)} \left| C_2(t^{-1}x, t^{-1}y) - R_2(x, y) \right| \to 0.$$

Assumption (2.a) is a standard second order condition in univariate extreme value theory; see, e.g. Section 2.3 in [10]. The condition $\gamma_1, \gamma_2 < \frac{1}{2}$ implies that there exists some $\delta_1 > 0$ such that $EX_+^{2+\delta_1} < \infty$ and $EY_+^{2+\delta_1} < \infty$. Assumption (2.b) can be viewed as a tail analogue of Assumption (1.b) and it is commonly assumed when applying the theory of tail copula process; see, e.g., [63], [64] and Theorem 7.2.2 in [10]. The R-function defined therein fully characterizes the so-called *stable tail dependence function* l in such a way that

$$l(x, y) = x + y - R(x, y), \quad x, y \geq 0;$$

see, e.g., [65] and Section 8.2 in [66].

Theorem 3.2.2 (Intermediate level). *Let $\alpha = \alpha_n$ be an intermediate sequence, that is, $\alpha_n \to 0$ and $n\alpha_n \to \infty$, as $n \to \infty$. Given $R(1,1) > 0$, Assumption 3.2.2 implies that, as $n \to \infty$,*

$$\sqrt{n\alpha}\left(\frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1\right) \xrightarrow{d} N\left(0, \sigma_0^2\right),$$

with $\sigma_0^2 = Var(\Lambda_0 + \Theta_{0,1} - \Theta_{0,2})$ and the zero-mean Gaussian random variables

$$\Lambda_0 = \frac{W_R(1,1) - R_1(1,1)W_R(1,\infty) - R_2(1,1)W_R(\infty,1)}{R(1,1)},$$
$$\Theta_{0,1} = (\gamma_1 - 1)\int_0^1 W_R(x,\infty)dx^{-\gamma_1}, \quad \Theta_{0,2} = (\gamma_2 - 1)\int_0^1 W_R(\infty,y)dy^{-\gamma_2}.$$

Here, W_R is a R-Brownian motion, i.e. a zero-mean Gaussian process with covariance function

$$\mathbb{E}(W_R(u_1,v_1)W_R(u_2,v_2)) = R(u_1 \wedge u_2, v_1 \wedge v_2) \quad \textit{for} \quad (u_1,v_1),(u_2,v_2) \in (0,\infty]^2 \backslash \{\infty,\infty\}.$$

Furthermore, if k is a symmetric density with support [-1,1] and bounded first derivative *and the bandwidth* $h = h(n) > 0$ *satisfies*

$$n\alpha h^2 \to \infty, \quad n\alpha h^4 \to 0, \quad \textit{and} \quad n\alpha h^2 A_i^2(1/\alpha) = O(1) \quad \textit{for} \quad i = 1, 2,$$

then we have that, as $n \to \infty$*,*

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha - \widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{P} 0.$$

Theorem 3.2.2 is a tail analogue of Theorem 3.2.1, despite slightly stronger conditions are imposed to eliminate the asymptotic bias due to the extreme-value approximations.

Based on these two asymptotic normality results, one can construct a confidence interval of ρ_α based on either $\widetilde{\rho}_\alpha$ or $\widehat{\rho}_\alpha$. Estimating the asymptotic variance of $\widetilde{\rho}_\alpha$ or $\widehat{\rho}_\alpha$ requires some (empirical) approximation of the copula function C or the function R, say, $\widehat{C}$ and $\widehat{R}$ respectively. A usual approach requires simulating the Gaussian process $B_{\widehat{C}}$ or $W_{\widehat{R}}$ with some empirical approximations of the limiting covariance functions. It is also necessary to estimate the first-order partial derivatives of the (tail) copula function and even, when α_n is an intermediate sequence, the tail indices of the marginal distributions. This approach is often quite computationally intensive, and its finite-sample performance can be quite poor by aggregating all the estimation uncertainties discussed above.

Instead, we investigate the possibility of employing the empirical likelihood method. Although this method proposed by [67] and [43] has proved to be quite effective in interval estimation and hypothesis testing, it has a serious problem in handling a nonlinear statistic. For example, it can lead to computational difficulties by solving a number (dependent on n) of simultaneous equations. Recently [60] proposed a so-called jackknife empirical likelihood method for dealing with nonlinear statistics such as U-statistics. Like inference for receiver operating characteristic (ROC) curves, copulas and tail copulas in [68], [59] and [58], a smoothed version is needed for the proposed relative risk measure.

Hence we shall establish our jackknife empirical likelihood inference method for ρ_α

based on the smoothed nonparametric estimation. To apply the smoothed jackknife empirical likelihood method, we first need to construct a jackknife pseudo sample of ρ_α given by

$$\widehat{V}_{\rho,i} = n\widehat{\rho}_\alpha - (n-1)\widehat{\rho}_{\alpha,i}, \quad i = 1, \ldots, n,$$

where

$$\widehat{\rho}_{\alpha,i} = \frac{1}{\alpha}\widehat{C}_i(\alpha,\alpha)\frac{\widehat{ES}_{\alpha,i}(X)}{\widehat{ES}_{\alpha,i}(Y)}$$

with

$$\begin{cases} \widehat{C}_i(\alpha,\alpha) = \frac{1}{n-1}\sum_{j\neq i} K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_j)/\alpha}{h}\right), \\ \widehat{ES}_{\alpha,i}(X) = \frac{1}{(n-1)\alpha}\sum_{j\neq i}(X_j - X_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) + X_{n-\lceil n\alpha\rceil:n}, \\ \widehat{ES}_{\alpha,i}(Y) = \frac{1}{(n-1)\alpha}\sum_{j\neq i}(Y_j - Y_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) + Y_{n-\lceil n\alpha\rceil:n}, \end{cases}$$

and

$$\bar{F}_{n1,i}(x) = \frac{1}{n-1}\sum_{j\neq i}\mathbb{1}\left[X_j > x\right], \quad \bar{F}_{n2,i}(y) = \frac{1}{n-1}\sum_{j\neq i}\mathbb{1}\left[Y_j > y\right], \quad x, y \in \mathbb{R}.$$

Based on this pseudo sample, the jackknife empirical likelihood ratio function for $\theta = \rho_\alpha$ can be defined by

$$\widehat{\mathcal{R}}(\theta) = \sup\left\{\Pi_{i=1}^n np_i : p_1 > 0, \ldots, p_n > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\widehat{V}_{\rho,i} = \theta\right\}.$$

Applying the Lagrange multiplier method yields

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(\widehat{V}_{\rho,i}-\theta)}, \tag{3.3}$$

where $\lambda = \lambda(\theta)$ solves the equation

$$\frac{1}{n}\sum_{i=1}^n \frac{\widehat{V}_{\rho,i}-\theta}{1+\lambda(\widehat{V}_{\rho,i}-\theta)} = 0. \tag{3.4}$$

It follows that the log empirical likelihood ratio is

$$-2\log\widehat{\mathcal{R}}(\theta) = 2\sum_{i=1}^{n}\log\left\{1+\lambda(\widehat{V}_{\rho,i}-\theta)\right\}.$$

To unify our jackknife empirical likelihood procedure for fixed and intermediate level α, we need one more assumption.

Assumption 3.2.3. For some $\tau > \max\{\gamma_1,\gamma_2\}$ such that

$$\lim_{t\downarrow 0}\sup_{0<x,y\leq 1}\frac{|t^{-1}C(tx,ty)-R(x,y)|}{(x\wedge y)^{\tau}} = 0, \tag{3.5}$$

where $x\wedge y := \min\{x,y\}$.

This is very similar to the condition (a) in [63] but we allow an arbitrary rate of convergence here. We can show that (3.5) is satisfied with $\tau = 1$ if

$$\lim_{t\downarrow 0}\sup_{x\geq 1}\left|t^{-1}C(tx,t)-R(x,1)\right| = 0 \text{ and } \lim_{t\downarrow 0}\sup_{y\geq 1}\left|t^{-1}C(t,ty)-R(1,y)\right| = 0,$$

which is weaker than the usual second-order condition (7.2.8) in [10].

Below is a Wilks type result for our JEL approach.

Theorem 3.2.3. *Either if the conditions of Theorem 3.2.1 hold, or if the conditions of Theorem 3.2.2 in conjunction with Assumption 3.2.3 hold, then $-2\log\widehat{\mathcal{R}}(\rho_\alpha)$ converges in distribution to a chi-squared limit with one degree of freedom as $n\to\infty$.*

Based on Theorem 3.2.3, an asymptotic confidence interval with level ψ for ρ_α is given by

$$I_\psi = \{\theta\in\mathbb{R} : -2\log\widehat{\mathcal{R}}(\theta)\leq \chi^2_{1,\psi}\}$$

where $\chi^2_{1,\psi}$ is the ψ-th quantile of the chi-squared distribution with one degree of freedom. This interval has the asymptotically correct coverage probability regardless of the level α being fixed or intermediate. In other words, for certain sample size n and small level α, both asymptotic embeddings lead to the same approximation. This interval can be efficiently determined using a standard search algorithm; for more details we refer to Section 2.9 in [1].

3.3 Relative Risk Measure for AR-GARCH Models

When observations follow from a strictly stationary sequence, it is a bit straightforward to derive the asymptotic limit of the proposed nonparametric estimator for the relative risk measure, while the construction of an interval generally relies on a blockwise bootstrap method or a blockwise empirical likelihood, which involves the difficult choice of a tuning parameter. Motivated by the considered real data analysis in Section 5, we consider the following AR-GARCH/IGARCH model

$$\begin{cases} X_t = \mu_x + \sum_{i=1}^{P_x} a_{x,i} X_{t-i} + \varepsilon_{x,t}, \\ \varepsilon_{x,t} = h_{x,t}^{1/2} \eta_{x,t}, \quad h_{x,t} = w_x + \sum_{i=1}^{q_x} \alpha_{x,i} \varepsilon_{x,t-i}^2 + \sum_{j=1}^{p_x} \beta_{x,j} h_{x,t-j} \end{cases} \tag{3.6}$$

and

$$\begin{cases} Y_t = \mu_y + \sum_{i=1}^{P_y} a_{y,i} Y_{t-i} + \varepsilon_{y,t}, \\ \varepsilon_{y,t} = h_{y,t}^{1/2} \eta_{y,t}, \quad h_{y,t} = w_y + \sum_{i=1}^{q_y} \alpha_{y,i} \varepsilon_{y,t-i}^2 + \sum_{j=1}^{p_y} \beta_{y,j} h_{y,t-j}, \end{cases} \tag{3.7}$$

where $(\eta_{x,t}, \eta_{y,t})'s$ are independent and identically distributed random vectors with means zero and variances one. In this case, we are concerned with both point and interval estimation of the relative risk measure of $\eta_{x,t}$ and $\eta_{y,t}$, which involves inference for unknown parameters in (3.6) and (3.7).

An obvious estimator for the unknown parameters in (3.6) and (3.7) is the so-called quasi maximum likelihood estimator, and its asymptotic normality is available in [69], which requires finite fourth moments of $\varepsilon_{x,t}, \eta_{x,y}, \varepsilon_{y,t}, \eta_{y,t}$. However, in practice it is quite often that $\sum_{i=1}^{q_x} \alpha_{x,i} + \sum_{j=1}^{p_x} \beta_{x,j}$ is close to one, and so assuming $E\varepsilon_{x,t}^4 < \infty$ may be problematic. Here we propose to employ the self-weighted estimator in [70] which requires weaker moment conditions when we use the weights

$$\delta_{x,t} = \left\{ \max(1, \frac{1}{C_x} \sum_{k=1}^{t-1} \frac{|X_{t-k}| I(|X_{t-k}| > C_x))}{k^9} \right\}^{-4}$$

and

$$\delta_{y,t} = \left\{ \max(1, \frac{1}{C_y} \sum_{k=1}^{t-1} \frac{|Y_{t-k}| I(|Y_{t-k}| > C_y))}{k^9} \right\}^{-4},$$

for some constants $C_x, C_y > 0$. These two constants are chosen as the 90% sample quantiles of $\{X_t\}_{t=1}^n$ and $\{Y_t\}_{t=1}^n$, respectively, as suggested by [71]. In the simulation study we use 95% sample quantiles of both samples, which performs better. Based on the self-weighted estimator in [70] with the above weight functions, we employ the corresponding score equations to formulate an empirical likelihood method to construct a confidence interval for the relative risk measure of $\eta_{x,t}$ and $\eta_{y,t}$ as follows.

Given the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ and the initial value $(\bar{X}_0, \bar{Y}_0) = \{(X_t, Y_t) : t \leq 0\}$ generated by the above model, we can write the parametric model as

$$\begin{gathered}
\varepsilon_{x,t}(\boldsymbol{\phi}_x) = X_t - \mu_x - \sum_{i=1}^{P_x} a_{x,i} X_{t-i}, \quad \varepsilon_{y,t}(\boldsymbol{\phi}_y) = Y_t - \mu_y - \sum_{i=1}^{P_y} a_{y,i} Y_{t-i}, \\
\eta_t(\boldsymbol{\psi}_x) = \varepsilon_t(\boldsymbol{\phi}_x) / \sqrt{h_t(\boldsymbol{\psi}_x)}, \quad \eta_t(\boldsymbol{\psi}_y) = \varepsilon_t(\boldsymbol{\phi}_y) / \sqrt{h_t(\boldsymbol{\psi}_y)}, \\
h_{x,t}(\boldsymbol{\psi}_x) = w_x + \sum_{i=1}^{q_x} \alpha_{x,i} \varepsilon_{x,t-i}^2(\boldsymbol{\phi}_x) + \sum_{i=1}^{p_x} \beta_{x,i} h_{x,t-i}(\boldsymbol{\psi}_x), \\
h_{y,t}(\boldsymbol{\psi}_y) = w_y + \sum_{i=1}^{q_y} \alpha_{y,i} \varepsilon_{y,t-i}^2(\boldsymbol{\phi}_y) + \sum_{i=1}^{p_y} \beta_{y,i} h_{y,t-i}(\boldsymbol{\psi}_y),
\end{gathered}$$

where $\boldsymbol{\phi}_x = (\mu_x, a_{x,1}, \ldots, a_{x,P_x})$, $\boldsymbol{\phi}_y = (\mu_y, a_{y,1}, \ldots, a_{y,P_y})$, $\boldsymbol{\phi}_{hx} = (w_x, \alpha_{x,1}, \ldots \alpha_{x,q_x}, \beta_{x,1}, \ldots, \beta_{x,p_x})$, $\boldsymbol{\phi}_{hy} = (w_y, \alpha_{y,1}, \ldots \alpha_{y,q_y}, \beta_{y,1}, \ldots, \beta_{y,p_y})$, $\boldsymbol{\psi}_x = (\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx})$, $\boldsymbol{\psi}_y = (\boldsymbol{\phi}_y, \boldsymbol{\phi}_{hy})$.

Then the self-weighted estimators in [70] for $\boldsymbol{\psi}_x^0$ and $\boldsymbol{\psi}_y^0$ solve the score equations

$$\sum_{t=1}^n \delta_{x,t}\frac{\partial l_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} := \sum_{t=1}^n \delta_{x,t}\frac{\partial}{\partial \boldsymbol{\psi}_x}\left\{\eta_{x,t}^2(\boldsymbol{\psi}_x)+\log h_{x,t}(\boldsymbol{\psi}_x)\right\}=0,$$
$$\sum_{t=1}^n \delta_{y,t}\frac{\partial l_{y,t}(\boldsymbol{\psi}_y)}{\partial \boldsymbol{\psi}_y} := \sum_{t=1}^n \delta_{y,t}\frac{\partial}{\partial \boldsymbol{\psi}_y}\left\{\eta_{y,t}^2(\boldsymbol{\psi}_y)+\log h_{x,t}(\boldsymbol{\psi}_y)\right\}=0,$$

respectively. Put the true parameters $\theta_1^0 = F_{\eta_{x,1}}^{\leftarrow}(1-\alpha)$, $\theta_2^0 = F_{\eta_{y,1}}^{\leftarrow}(1-\alpha)$, $\theta_3^0 = E\{\eta_{x,1}\mathbf{1}(\eta_{x,1} > \theta_1^0)\}$, $\theta_4^0 = E\{\eta_{y,1}\mathbf{1}(\eta_{y,1} > \theta_2^0)\}$, $\theta_5^0 = \rho_\alpha(\eta_{x,1},\eta_{y,1})$. Then $F_{\eta_{x,1}}(\theta_1^0) = 1-\alpha$, $F_{\eta_{y,1}}(\theta_2^0) = 1-\alpha$, $\theta_3^0 = -E\{\eta_{x,1}\mathbf{1}(\eta_{x,1} \le \theta_1^0)$, $\theta_4^0 = -E\{\eta_{y,1}\mathbf{1}(\eta_{y,1} \le \theta_2^0)$, $\theta_5^0\alpha\theta_4^0/\theta_3^0 = P(\eta_{x,1} > \theta_1^0, \eta_{y,1} > \theta_2^0) = 2\alpha - 1 + P(\eta_{x,1} \le \theta_1^0, \eta_{y,1} \le \theta_2^0)$, which motivate the following definitions

$$\begin{aligned}\boldsymbol{Z}_t(\theta_5,\boldsymbol{\nu}) =&\Bigg(\delta_{x,t}\frac{\partial l_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}, K\left(\frac{\theta_1-\eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right)-1+\alpha, -\eta_{x,t}(\boldsymbol{\psi}_x)K\left(\frac{\theta_1-\eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right)-\theta_3,\\ &\delta_{y,t}\frac{\partial l_{y,t}(\boldsymbol{\psi}_y)}{\partial \boldsymbol{\psi}_y}, K\left(\frac{\theta_2-\eta_{y,t}(\boldsymbol{\psi}_y)}{h}\right)-1+\alpha, -\eta_{y,t}(\boldsymbol{\psi}_y)K\left(\frac{\theta_2-\eta_{y,t}(\boldsymbol{\psi}_y)}{h}\right)-\theta_4,\\ &2\alpha-1+K\left(\frac{\theta_1-\eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right)K\left(\frac{\theta_2-\eta_{y,t}(\boldsymbol{\psi}_x)}{h}\right)-\alpha\theta_5\theta_4/\theta_3\Bigg)\end{aligned} \tag{3.8}$$

where $\boldsymbol{\nu} := (\boldsymbol{\psi}_x, \boldsymbol{\psi}_y, \theta_1, \ldots, \theta_4)$.

Then the empirical likelihood function for θ_5 and $\boldsymbol{\nu}$ is defined as

$$L(\theta_5,\boldsymbol{\nu}) = \sup\{\prod_{t=1}^n (nr_t) : r_1 \ge 0, \cdots, r_n \ge 0, \sum_{t=1}^n r_t = 1, \sum_{t=1}^n r_t \boldsymbol{Z}_t(\theta_5,\boldsymbol{\nu}) = 0\}.$$

Since we are only interested in $\theta_5^0 = \rho_\alpha(\eta_{x,1},\eta_{y,1})$, we consider the profile empirical likelihood function

$$L^P(\theta_5) = \max_{\boldsymbol{\nu}} L(\theta_5,\boldsymbol{\nu}).$$

To show that Wilks theorem holds for the above proposed profile empirical likelihood method, we assume the following regularity conditions:

- A1) Let Θ_x and Θ_y denote the parameter spaces for $\boldsymbol{\psi}_x$ and $\boldsymbol{\psi}_y$, separately. Assume the true values of $\boldsymbol{\psi}_x$ and $\boldsymbol{\psi}_y$ are an interior of Θ_x and Θ_y, respectively;

- A2) $1 - \sum_{i=1}^{P_x} a_{x,i} z^i \neq 0$ for all $|z| \leq 1$ and $\boldsymbol{\psi}_x \in \Theta_x$, and $1 - \sum_{i=1}^{P_y} a_{y,i} z^i \neq 0$ for all $|z| \leq 1$ and $\boldsymbol{\psi}_y \in \Theta_y$;

- A3) There is no common root for equations $\sum_{i=1}^{q_x} \alpha_{x,i} z^i = 0$ and $1 - \sum_{j=1}^{p_x} \beta_{x,j} z^j = 0$, $\sum_{i=1}^{q_x} \alpha_{x,i} \neq 0$, $\alpha_{x,q_x} + \beta_{x,p_x} \neq 0$ and $\sum_{j=1}^{p_x} \beta_{x,j} < 1$ for each $\boldsymbol{\psi}_x \in \Theta_x$. Similarly there is no common root for equations $\sum_{i=1}^{q_y} \alpha_{y,i} z^i = 0$ and $1 - \sum_{j=1}^{p_y} \beta_{y,j} z^j = 0$, $\sum_{i=1}^{q_y} \alpha_{y,i} \neq 0$, $\alpha_{y,q_y} + \beta_{y,p_y} \neq 0$ and $\sum_{j=1}^{p_y} \beta_{y,j} < 1$; for each $\boldsymbol{\psi}_y \in \Theta_y$.

- A4) $\{(\eta_{x,t}, \eta_{y,t})^T\}_{t=1}^n$ is a sequence of independent and identically distributed random vectors satisfying $E\eta_{x,t} = 0$, $E\eta_{x,t}^2 = 1$, $E|\eta_{x,t}|^{4+\delta_0} < \infty$, $E\eta_{y,t} = 0$, $E\eta_{y,t}^2 = 1$ and $E|\eta_{y,t}|^{4+\delta_0} < \infty$ for some $\delta_0 > 0$. $\eta_{x,t}$ and $\eta_{y,t}$ are not perfectly correlated.

- A5) $\sum_{i=1}^{q_x} \alpha_{x,i} + \sum_{i=1}^{p_x} \beta_{x,i} \leq 1$ and $\sum_{i=1}^{q_y} \alpha_{y,i} + \sum_{i=1}^{p_y} \beta_{y,i} \leq 1$. Moreover, $\eta_{x,t}$ or/and $\eta_{y,t}$ have positive density on $\mathbb{R}$ when $\sum_{i=1}^{q_x} \alpha_{x,i} + \sum_{i=1}^{p_x} \beta_{x,i} = 1$ or/and $\sum_{i=1}^{q_y} \alpha_{y,i} + \sum_{i=1}^{p_y} \beta_{y,i} = 1$.

- A6) Assumption 1 in Theorem 3.2.1 holds for $\eta_{x,t}$ and $\eta_{y,t}$, and other conditions in Theorem 3.2.1 are true too. Furthermore, $nh^{2+\delta_h} \to \infty$ for some $\delta_h > 0$.

Assumptions A1)–A3) are standard conditions for the stationarity and identifiability of the AR-GARCH/IGARCH model. Assumption A4) implies finite fourth moments of $\eta_{x,t}$ and $\eta_{y,t}$, which are necessary for the asymptotic normality of self-weighted estimator. Assumption A5) implies that $E(|\varepsilon_t|^{2\iota}) < \infty$ for all $\iota \in (0, 1)$. See, e.g., Sections 2 and 3 in [70] for more discussions.

Note that we focus on the case of a fixed level α. When $\alpha \to 0$, the first step in estimating unknown parameters using self-weighted estimator or local QMLE from [70] in models (3.6) and (3.7) does not play a role asymptotically in estimating the relative risk measure of $\eta_{x,t}$ and $\eta_{y,t}$ due to their fast rate of convergence under suitable conditions. In other words, Theorem 3.2.3 is still applicable to models (3.6) and (3.7) in case of an intermediate level under Assumptions A1)–A5) and the conditions in Theorem 3.2.2 on $\eta_{x,t}$ and $\eta_{y,t}$.

Theorem 3.3.1. *Assume models (3.6) and (3.7) hold for conditions A1)–A6). Then, as $n \to \infty$, $-2 \log L^P(\rho_\alpha(\eta_{x,1}, \eta_{y,1}))$ converges in distribution to a chi-squared limit with one degree of freedom.*

As before, an empirical likelihood confidence interval for the relative risk measure $\rho_\alpha(\eta_{x,1}, \eta_{y,1})$ based on models (3.6) and (3.7) can be obtained via the above theorem.

Remark 3.3.1. Similarly under models (3.6) and (3.7) we can develop a profile empirical likelihood interval for the conditional relative risk measure of (X_{n+1}, Y_{n+1}) given $(X_1, Y_1), \cdots, (X_n, Y_n)$ and $h_{x,n}, \cdots, h_{x,n-p_x}, h_{y,n}, \cdots, h_{y,n-p_y}$ by noting that $h_{x,n+1}$ and $h_{y,n+1}$ can be expressed as functions of conditional variables and unknown parameters in the above AR-GARCH models.

3.4 Simulation Study

3.4.1 Independent data

In this subsection, a simulation study is carried out to evaluate the finite-sample behavior of the proposed jackknife empirical likelihood method for our proposed relative risk measure ρ_α. The survival copula in our simulation study is a so-called t-copula with multiple parameters of degrees of freedom which is a generalization of the grouped t-copula; see [72] for details. The distribution of a two-dimensional t-copula with multiple parameters of degrees of freedom is

$$C^{\Sigma}_{\nu_1,\nu_2}(u_1, u_2) = \int_0^1 \Phi_\Sigma(z_1(u_1, s), z_2(u_2, s))ds, \quad u_1, u_2 \in [0, 1],$$

- Φ_Σ is the distribution function of a bivariate normal random vector with zero means, unit variances and positive correlation ρ;

- $z_i(u_i, s) = t_{\nu_i}^{-1}(u_i)/\omega_i(s), \omega_i(s) = \sqrt{\nu_i/\chi_{\nu_i}^{-1}(s)}$, $i = 1, 2$;

- t_{ν_i} and $t_{\nu_i}^{-1}$ denote the distribution function and quantile of a student-t random variable with ν_i degrees of freedom respectively, $i = 1, 2$;

- χ_{ν_i} and $\chi_{\nu_i}^{-1}$ denote the distribution function and quantile of a chi-squared random variable with ν_i degrees of freedom respectively, $i = 1, 2$.

We draw 1000 random samples of size $n = 500$ and 1000 from a bivariate distribution with a t-copula with two parameters of degrees of freedom $\nu = (\nu_1, \nu_2) \in \{(3,3), (3,5), (5,3), (5,5)\}$ and two marginal t distributions with degrees of freedom ν_1 and ν_2 respectively. We consider two cut-off levels $\alpha = 0.05, 0.1$ and two confidence levels $\psi = 0.9, 0.95$. In all cases, we set $\rho = 0.2$.

Table 3.1: Empirical coverage probabilities for the jackknife empirical likelihood-based confidence interval $I_\psi(h)$ and the bootstrap confidence interval I_ψ^* of ρ_α with cutoff level $\boldsymbol{\alpha = 0.05}$, sample size $n = 500, 1000$ and confidence levels $\psi = 0.90, 0.95$. Bandwidths are chosen as $h_1 = 0.5(n\alpha)^{-\frac{1}{3}}$, $h_2 = (n\alpha)^{-\frac{1}{3}}$, $h_3 = 1.5(n\alpha)^{-\frac{1}{3}}$, $h_4 = 2(n\alpha)^{-\frac{1}{3}}$, $h_5 = 2.5(n\alpha)^{-\frac{1}{3}}$, $h_6 = 3(n\alpha)^{-\frac{1}{3}}$.

	$n = 500$				$n = 1000$			
(ν_1, ν_2)	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$
$I^*_{0.95}$	0.880	0.873	0.882	0.860	0.920	0.922	0.914	0.919
$I_{0.95}(h_1)$	0.938	0.954	0.939	0.945	0.934	0.934	0.937	0.943
$I_{0.95}(h_2)$	0.947	0.959	0.953	0.960	0.947	0.945	0.951	0.939
$I_{0.95}(h_3)$	0.945	0.954	0.956	0.958	0.945	0.940	0.955	0.946
$I_{0.95}(h_4)$	0.940	0.950	0.952	0.955	0.932	0.932	0.935	0.939
$I_{0.95}(h_5)$	0.941	0.944	0.949	0.955	0.927	0.934	0.933	0.935
$I_{0.95}(h_6)$	0.925	0.938	0.947	0.950	0.923	0.924	0.927	0.932
$I^*_{0.90}$	0.834	0.824	0.838	0.831	0.868	0.870	0.874	0.867
$I_{0.90}(h_1)$	0.890	0.903	0.886	0.889	0.883	0.897	0.892	0.890
$I_{0.90}(h_2)$	0.900	0.914	0.911	0.915	0.899	0.892	0.890	0.900
$I_{0.90}(h_3)$	0.896	0.903	0.903	0.910	0.902	0.893	0.899	0.898
$I_{0.90}(h_4)$	0.885	0.899	0.905	0.900	0.884	0.881	0.885	0.889
$I_{0.90}(h_5)$	0.880	0.892	0.901	0.903	0.879	0.876	0.887	0.890
$I_{0.90}(h_6)$	0.865	0.882	0.885	0.896	0.875	0.873	0.876	0.889

The empirical coverage probability of the jackknife empirical likelihood-based confidence interval is compared to that of the bootstrap confidence interval. The bootstrap confidence interval is obtained by using 1000 bootstrap samples of size n from each sample $X_1, \ldots, X_n$. Specifically, for each bootstrap sample, we calculate the empirical estimate of ρ_α, which results in 1000 bootstrapped empirical estimates of ρ_α, denoted as $\widetilde{\rho}_\alpha^{*1}, \ldots, \widetilde{\rho}_\alpha^{*1000}$, and therefore 1000 bootstrap differences $\delta^{*i} = \widetilde{\rho}_\alpha^{*i} - \rho_\alpha$, $i = 1, \ldots, 1000$. Ordering these bootstrap differences by $\delta^{*[1]} \leq \ldots \leq \delta^{*[1000]}$, the bootstrap confidence interval at level ψ is then calculated as

$$I_\psi^* = [\widetilde{\rho}_\alpha - \delta^{*[n2]},\ \widetilde{\rho}_\alpha - \delta^{*[n1]}],$$

where n_1 and n_2 denote the integer part of $500(1-\psi)$ and $500(1+\psi)$, respectively. Motivated by the optimal bandwidth choice in smoothing distribution function estimation, we choose $h = d(n\alpha)^{-1/3}$ for various $d = 0.5, 1, 1.5, 2, 2.5, 3$.

We report the empirical coverage probabilities in Tables 3.1 and 3.2, which show that the proposed jackknife empirical likelihood method performs better than the bootstrap method in terms of coverage accuracy, and the results are quite stable with respect to the different choices of bandwidth h especially with $d = 1, 1.5, 2$. For $\alpha = 0.1$, it clearly shows that a larger size improves the accuracy.

3.4.2 AR-GARCH Models

In addition to the independent cases, we also implement the simulation studies for models (3.6) and (3.7) by using similar settings to the real data in Section 3.5. In real data, we multiply observations (loss) by 100 for proper scaling, and then fit the models (3.6) and (3.7) which turn out AR(1)+GARCH(1,1) fits well for both margins by checking the autocorrelation functions of estimated residuals. Note that we do not use AIC or BIC to choose the best fitting and simply prefer AR-GARCH models with a smaller number of parameters due to the heavy computation in the profile empirical likelihood method. We also find that the innovations in models (3.6) and (3.7) are fitted well by marginal t distributions with

Table 3.2: Empirical coverage probabilities for the jackknife empirical likelihood-based confidence interval $I_\psi(h)$ and the bootstrap confidence interval I_ψ^* of ρ_α with cutoff levels $\boldsymbol{\alpha = 0.1}$, sample size $n = 500, 1000$ and confidence levels $\psi = 0.90, 0.95$. Bandwidths are chosen as $h_1 = 0.5(n\alpha)^{-\frac{1}{3}}$, $h_2 = (n\alpha)^{-\frac{1}{3}}$, $h_3 = 1.5(n\alpha)^{-\frac{1}{3}}$, $h_4 = 2(n\alpha)^{-\frac{1}{3}}$, $h_5 = 2.5(n\alpha)^{-\frac{1}{3}}$, $h_6 = 3(n\alpha)^{-\frac{1}{3}}$.

	$n = 500$				$n = 1000$			
(ν_1, ν_2)	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$
$I_{0.95}^*$	0.920	0.922	0.914	0.919	0.935	0.939	0.937	0.935
$I_{0.95}(h_1)$	0.930	0.940	0.938	0.934	0.950	0.943	0.948	0.953
$I_{0.95}(h_2)$	0.938	0.946	0.947	0.950	0.948	0.946	0.954	0.959
$I_{0.95}(h_3)$	0.942	0.949	0.944	0.947	0.946	0.947	0.950	0.956
$I_{0.95}(h_4)$	0.932	0.932	0.935	0.939	0.951	0.952	0.943	0.946
$I_{0.95}(h_5)$	0.927	0.934	0.933	0.935	0.948	0.947	0.937	0.946
$I_{0.95}(h_6)$	0.923	0.924	0.927	0.932	0.943	0.947	0.939	0.940
$I_{0.90}^*$	0.868	0.870	0.874	0.867	0.874	0.887	0.874	0.890
$I_{0.90}(h_1)$	0.880	0.885	0.877	0.882	0.911	0.890	0.906	0.905
$I_{0.90}(h_2)$	0.884	0.891	0.884	0.891	0.909	0.906	0.901	0.904
$I_{0.90}(h_3)$	0.879	0.894	0.888	0.891	0.910	0.903	0.897	0.907
$I_{0.90}(h_4)$	0.884	0.881	0.885	0.889	0.897	0.903	0.893	0.896
$I_{0.90}(h_5)$	0.879	0.876	0.887	0.890	0.887	0.899	0.891	0.899
$I_{0.90}(h_6)$	0.875	0.873	0.876	0.889	0.880	0.892	0.886	0.893

degrees of freedom 5.5, 8.9 and a t-copula in real analysis. Therefore, we simulate data from the following AR(1)-GARCH(1,1) models

$$X_t = \mu_x + a_x X_{t-1} + \varepsilon_{x,t}, \quad \varepsilon_{x,t} = h_{x,t}^{1/2}\eta_{x,t}, \quad h_{x,t} = \omega_x + \alpha_x \varepsilon_{x,t-1}^2 + \beta_x h_{x,t-1},$$

$$Y_t = \mu_y + a_y Y_{t-1} + \varepsilon_{y,t}, \quad \varepsilon_{y,t} = h_{y,t}^{1/2}\eta_{y,t}, \quad h_{y,t} = \omega_y + \alpha_y \varepsilon_{y,t-1}^2 + \beta_y h_{y,t-1},$$

where the i.i.d innovations $\{(\eta_{x,t}, \eta_{y,t})\}_{t=1}^n$ have marginal t-distribution with degrees of freedom 5.5, 8.9 scaled to unit variance and a t-copula with correlation γ and common degree of freedom 7. Details of the parameter setting from real data can be found in Section 3.5.

The results of eight simulation cases are summarized in Table 3.3, which shows the proposed method works well for the considered time series models, and accuracy improves as the sample size becomes larger.

Table 3.3: Empirical coverage probabilities for the jackknife empirical likelihood-based confidence interval $I_\psi(h)$ of ρ_α with cutoff levels $\boldsymbol{\alpha = 0.05}$, sample size $n = 2000$ and 4000, and confidence level $\psi = 0.95$. Bandwidths are chosen as $h_1 = 0.5(n\alpha)^{-\frac{1}{3}}$, $h_2 = (n\alpha)^{-\frac{1}{3}}$, $h_3 = 1.5(n\alpha)^{-\frac{1}{3}}$. The parameters in AR(1)-GARCH(1,1) are $\mu_x = -0.030, \mu_y = -0.053, \omega_x = -0.029, \omega_y = 0.012, \alpha_x = 0.072, \alpha_y = 0.074$.

	$I_{0.95}\&n = 2000$			$I_{0.95}\&n = 4000$		
$(\beta_x, \beta_y, \gamma)$	h_1	h_2	h_3	h_1	h_2	h_3
$(0.924, 0.917, 0.72)$	0.941	0.932	0.934	0.936	0.940	0.928
$(0.824, 0.817, 0.72)$	0.933	0.931	0.923	0.941	0.933	0.930
$(0.924, 0.917, 0)$	0.969	0.968	0.966	0.958	0.951	0.951
$(0.824, 0.817, 0)$	0.967	0.971	0.967	0.955	0.943	0.943

3.5 Real-life Data Analysis

In this section, we study our relative risk measure in a real-life data set, which contains daily stock losses on three U.S. Banks, Bank of America (BOA), JP Morgan (JPM), Wells Fargo (WFG), and Standard & Poor's 500 index (SNP) (benchmark) between February 1st, 2002 and March 31st, 2011 from the Center for Research in Security Prices (CRSP), and (weekly) levels of the Adjusted National Financial Conditional Index (ANFCI) between September 1st, 2006 and March 30, 2011. As usual, our collected stock returns exhibit so-called *volatility clustering* behavior widely documented in the empirical finance literature: the univariate squared stock returns are moderately auto-correlated. Hence, we shall work on a filtered version of the univariate losses. For loss data of each bank combined with S&P500 index we calibrate an autoregressive moving average model ARMA(p,q) with generalized autoregressive conditional heteroskedasticity (proposed in [73]) GARCH(P,Q) errors. Due to the heavy computation in the proposed profile empirical likelihood method, we prefer ARMA-GARCH models with a smaller number of parameters as long as the estimated residuals are uncorrelated. Hence we use the following AR(1)-GARCH(1,1) models:

$$X_t = \mu_x + a_x X_{t-1} + \varepsilon_{x,t}, \quad \varepsilon_{x,t} = h_{x,t}^{1/2}\eta_{x,t}, \quad h_{x,t} = \omega_x + \alpha_x \varepsilon_{x,t-1}^2 + \beta_x h_{x,t-1},$$

$$Y_t = \mu_y + a_y Y_{t-1} + \varepsilon_{y,t}, \quad \varepsilon_{y,t} = h_{y,t}^{1/2}\eta_{y,t}, \quad h_{y,t} = \omega_y + \alpha_y \varepsilon_{y,t-1}^2 + \beta_y h_{y,t-1},$$

where X_t is the stock loss on an individual institution and Y_t is the benchmark (loss on S&P 500 in our study).

After estimating the parameters in the above AR-GARCH models by the self-weighted QMLE in [70], we plot the autocorrelation functions of the estimated residuals in Figure 3.1, which shows the assumption of independent $(\eta_{x,t}, \eta_{y,t})'s$ is reasonable. That is, the proposed profile empirical likelihood method is applicable.

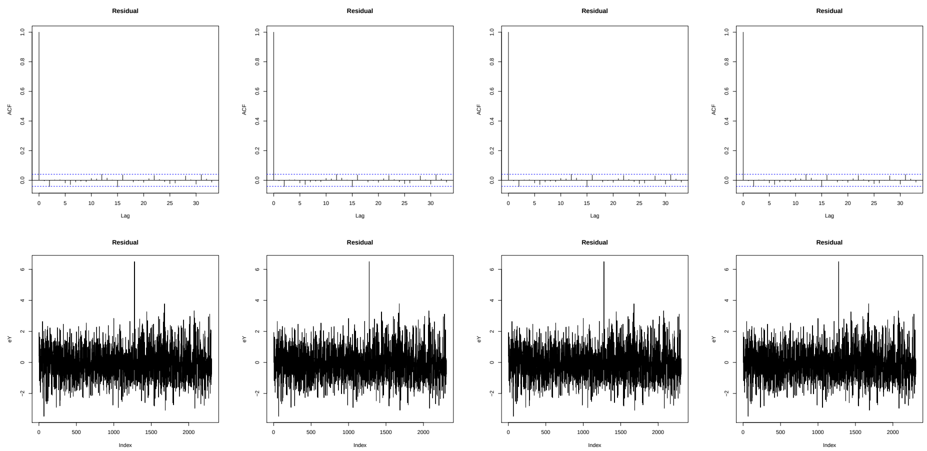
Table 3.4: Estimated coefficients of AR(1)-GARCH(1,1) models for three banks and S&P500 index.

	df(Marginal)	γ	df(tcopula)	μ	a	ω	α	β
BOA	5.5	0.72	7	-0.021	-0.029	0.020	0.067	0.929
JPM	6.3	0.77	7	-0.064	-0.037	0.014	0.076	0.923
WFG	6.5	0.72	7	-0.055	-0.104	0.011	0.099	0.900
SNP	8.9	N/A	N/A	-0.052	-0.083	0.012	0.074	0.917

Since we use the same setting to generate data to evaluate the finite sample performance of the proposed method in the simulation study, we need a parametric model for $(\eta_{x,t}, \eta_{y,t})^T$. Here we fit innovations $\{(\eta_{x,t}, \eta_{y,t})\}_{t=1}^n$ by marginal t-distributions with degrees of freedom df_x, df_y scaled to unit variance and a t-copula with correlation γ and common degree of freedom df. The fitted results are summarized in Table 3.4. Note that $\eta_{x,t}$ is related to either BOA or JPM or WFG, and $\eta_{y,t}$ is related to SNP.

Given $(X_1, Y_1), \cdots, (X_t, Y_t)$, the relative risk measure of (X_{t+1}, Y_{t+1}) can not be written as the relative risk measure of $(\eta_{x,t+1}, \eta_{y,t+1})$ multiplied by a constant, but the relative risk measure of $(\varepsilon_{t+1}, \varepsilon_{t+1})$ can be written as the relative risk measure of $(\eta_{x,t+1}, \eta_{y,t+1})$ multiplied by the constant $h_{x,t}^{1/2}/h_{y,t}^{1/2}$. With a focus on volatility, we estimate the relative

Figure 3.1: The columns from left to right are for Bank of America, JP Morgan, Wells Fargo and S&P500. The first row are ACF plots from the AR(1)+GARCH(1,1) models of three banks and S&P500, and the second row are residual plots.



risk measure of $(\varepsilon_{x,t+1}, \varepsilon_{y,t+1})$ given (X_s, Y_s) for $s \leq t$, i.e.,

$$\rho_{\alpha,t+1|t} = \sqrt{\frac{h_{x,t+1}}{h_{y,t+1}}} \rho_\alpha(\eta_{x,1}, \eta_{y,1}), \tag{3.9}$$

where $\rho_\alpha(\eta_{x,1}, \eta_{y,1})$ is the (unconditional) relative risk measure of $(\eta_{x,1}, \eta_{y,1})$.

We plot the estimated $\rho_{\alpha,t+1|t}$ for each of the three banks against S&P500 index in Figure 3.2 from February 1st, 2002 to March 31st, 2011. The intervals, at 95% level, are constructed by firstly the profile empirical likelihood method on the estimated innovations and then a simple multiplication of the estimated $\sqrt{\frac{h_{x,t+1}}{h_{y,t+1}}}$ at time $t+1$. Note that we ignore the uncertainty of estimating $h_{x,t+1}$ and $h_{y,t+1}$ in the above intervals. One can easily spot some features of the three time series in Figure 3.2. First, during the pre-crisis period (2002-2007), all three banks have relative risk measures below the market level, i.e. unit level, for most of the time where Wells Fargo has the lowest variation (in the sense of shortest interval) and JP Morgan has the highest one. Second, during the crisis period (2008-2010), all three banks share a similar pattern of the relative risk measures, which grow suddenly to a peak value and then varies between high risk values. However, there is no obvious similar pattern during the non-crisis for the three banks. This suggests that the three banks may encounter (substantial) systemic risk during the crisis period. We observe that the average relative risk level of Bank of American is (more than 20%) higher than that of Wells Fargo and JP Morgan during April 2008 to March 2009, which is consistent with the ranking using sample marginal expected shortfalls in the same period; see Table 1 in [74].

Next, we document some empirical evidence of the predictive power of bank-specific relative risk on the distribution of future systemic shocks during the crisis period. We measure systemic shocks by the innovations to an autoregression to ANFCI. Using a rolling window size of 1154 days (half of our sample size), we construct the daily prediction of relative risks in a recursive manner using (3.9) where $\rho_\alpha(\eta_{x,1}, \eta_{y,1})$ is estimated by maximizing the profile empirical likelihood $L^P(\theta_5)$ defined in Section 3 and the AR(1)-GARCH(1,1) parameters are estimated by the self-weighted estimator proposed in [70]. Since the ANFCI is only reported weekly (every Friday), our predicted relative risks are converted into a weakly basis by taking the median of the daily estimates in the same week. Given some forecast horizon $L > 0$, forecast accuracy can be evaluated via an out-of-sample R^2 based on the mean squared loss function:

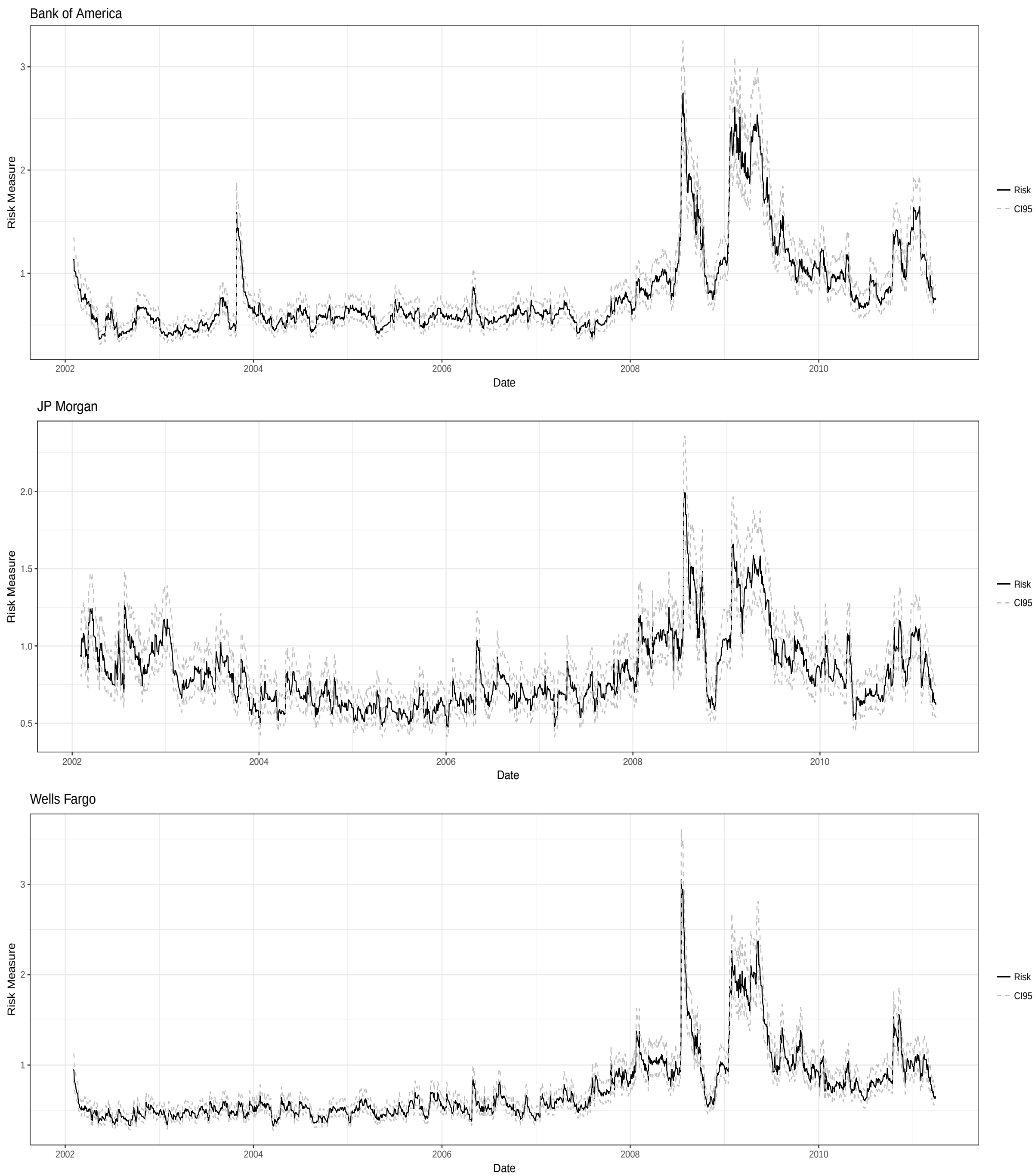
$$R^2 = 1 - \frac{\sum_t (ANFCI_t - \widehat{ANFCI}_t)^2}{\sum_t (ANFCI_t - \overline{ANFCI}_t)^2} \tag{3.10}$$

where $\widehat{ANFCI}_t$ is the fitted value from a predictive regression estimated through period $t-1$ using both the L-weeks lagged ANFCI and L-weeks lagged individual relative risk, and $\overline{ANFCI}_t$ is that from the corresponding autoregression using L-weeks lagged ANFCI only. The models including more lagged variables of ANFCI produce qualitatively the same results in our analysis.

Table 3.5: Out-of-sample R^2 (in percentage) for the predictive regressions with bank-specific relative risks in forecast horizon $L = 8, 9, 10$ weeks.

R^2(%)	BOA	JPM	WFG
$L = 8$	-2.14	1.74	2.31
$L = 9$	-0.62	2.98	5.19
$L = 10$	1.07	2.52	4.80

Figure 3.2: Time Series from February 1st, 2002 to March 31st, 2011 and 95% intervals of the relative tail risk measure in (3.9) with cut-off level $\alpha = 0.05$ for each of the daily stock losses on J.P. Morgan, Bank of America and Wells Fargo against the daily loss on Standard & Poor's 500 index. $h = (n\alpha)^{-1/3}$.



We forecast *ANFCI* starting from December 2007 in $L = 8, 9, 10$ weeks (about two months) forecast horizon and the out-of-sample R^2 are reported in Table 3.5. The predictive power of the three banks seems to be quite different, although their estimated relative risk are highly correlated over time. We observe positive out-of-sample R^2's for JP Morgan and Wells Fargo, which means the predictive regression including their individual relative risk has lower average mean squared prediction error than that of the corresponding autoregressive model. The estimated relative risk of BOA, in contrast, has shown almost no predictive power (beyond the autoregression).

In summary, the proposed relative risk measure and its intervals are useful in monitoring systemic risk.

3.6 Proofs

3.6.1 Independent Data

This section starts from the asymptotics of $\widehat{\rho}_\alpha$ and $\widetilde{\rho}_\alpha$, i.e. Theorems 3.2.1 and 3.2.2. Our proofs will be based on some (well-known) asymptotic results of the weighted empirical copula process (e.g., Appendix G in [75]) and their tail analogues (see, e.g., Lemma 1 in [63]). Define the *pseudo* estimator of the survival copula function by

$$\check{C}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[\bar{F}_1(X_i) < x, \bar{F}_2(Y_i) < y\right] \quad \text{for} \quad (x,y) \in [0,1]^2,$$

and, accordingly, the *pseudo* tail copula empirical process

$$\begin{aligned}\mathbb{W}_n(x,y) &= \sqrt{n\alpha}\left(\frac{1}{\alpha}\check{C}_n(\alpha x, \alpha y) - \frac{1}{\alpha}C(\alpha x, \alpha y)\right)\\ &=: \sqrt{n\alpha}\left(\check{R}_n(x,y) - R_n(x,y)\right) \quad \text{for} \quad (x,y) \in (0,\infty]^2 \setminus \{\infty,\infty\}.\end{aligned}$$

Below '$\xrightarrow{w}$' denotes weak convergence, $D(I)$ denotes the Skorohod space defined on domain I. Recall that '$\wedge$' denotes the minimum operator and '$\xrightarrow{P}$' denotes convergence in probability.

Lemma 3.6.1. *Suppose Assumption 3.2.1 hold and introduce a weighting function* $q_\eta(t) :=$

$t^\eta(1-t)^\eta$, $t>0$, *with* $\eta \in [0,1/2)$. *For a fixed* $\alpha \in (0,1)$, *as* $n\to\infty$, *in* $D([0,1/\alpha]^2)$

$$\left\{\frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x \wedge \alpha y)},\ x,y\in[0,1/\alpha]\right\} \xrightarrow{w} \left\{\frac{1}{\sqrt{\alpha}}\frac{B_C(\alpha x,\alpha y)}{q_\eta(\alpha x\wedge \alpha y)},\ x,y\in[0,1/\alpha]\right\},$$

where we shall read $\frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x\wedge\alpha y)}=0$ *and* $\frac{B_C(\alpha x,\alpha y)}{q_\eta(\alpha x\wedge\alpha y)}=0$ *if* $x=0$ *or* $y=0$ *or* $x=y=1/\alpha$.

Proof. See Proposition G.1 in [75]. □

Lemma 3.6.2. *Let* α *be an intermediate sequence such that* $\alpha=\alpha_n\to 0$ *and* $n\alpha_n\to\infty$ *and suppose condition* (3.2) *holds. For any* $\eta\in[0,1/2)$ *and* T *positive, as* $n\to\infty$,

$$\begin{aligned}&\left(\frac{\mathbb{W}_n(x,y)}{(x\wedge y)^\eta},(x,y)\in(0,T]^2,\frac{\mathbb{W}_n(x,\infty)}{x^\eta},x\in(0,T],\frac{\mathbb{W}_n(\infty,y)}{y^\eta},y\in(0,T]\right)\\ &\qquad\xrightarrow{w}\left(\frac{W_R(x,y)}{(x\wedge y)^\eta},(x,y)\in(0,T]^2,\frac{W_R(x,\infty)}{x^\eta},x\in(0,T],\frac{W_R(\infty,y)}{y^\eta},y\in(0,T]\right)\end{aligned}$$

in $D((0,T]^2)\times D((0,T])\times D((0,T])$.

Proof. For convenient presentation, all the limit processes below are defined on the same probability space, via the Skorohod construction. From Lemma 1 in [63] we know that

$$\begin{aligned}&\left(\frac{\mathbb{W}_n(x,y)}{x^\eta},(x,y)\in(0,T]^2,\frac{\mathbb{W}_n(x,\infty)}{x^\eta},x\in(0,T],\frac{\mathbb{W}_n(\infty,y)}{y^\eta},y\in(0,T]\right)\\ &\qquad\xrightarrow{a.s.}\left(\frac{W_R(x,y)}{x^\eta},(x,y)\in(0,T]^2,\frac{W_R(x,\infty)}{x^\eta},x\in(0,T],\frac{W_R(\infty,y)}{y^\eta},y\in(0,T]\right).\end{aligned}$$

Similarly as that in the first coordinate above, we can also show that

$$\left(\frac{\mathbb{W}_n(x,y)}{y^\eta},(x,y)\in(0,T]^2\right)\xrightarrow{a.s.}\left(\frac{W_R(x,y)}{y^\eta},(x,y)\in(0,T]^2\right).$$

Hence,

$$\sup_{x,y\in(0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{(x\wedge y)^\eta} = \max\left\{ \sup_{x,y\in(0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{x^\eta}, \right.$$
$$\left. \sup_{x,y\in(0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{y^\eta} \right\}$$
$$\xrightarrow{a.s.} \max\{0,0\} = 0.$$

The claim then follows. □

Lemma 3.6.3. *Under the conditions of Theorem 3.2.1 or Theorem 3.2.2, as* $n \to \infty$*,*

$$\frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} = o_P((n\alpha h^2)^{-1/2}),$$
$$\frac{Y_{n-[n\alpha(1-h))]:n} - Y_{n-[n\alpha(1+h)]:n}}{Q_2(1-\alpha)} = o_P((n\alpha h^2)^{-1/2}).$$

Proof. We only prove the first statement since the proof of the second one is completely analogous. For the fixed $\alpha \in (0,1)$, by the classical theory of quantile process (c.f., e.g., Example V.12 in [76]) and Assumption (1.a), we have that, for large n,

$$\begin{aligned} X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n} &= Q_1(1-\alpha(1-h)) - Q_1(1-\alpha(1+h)) + O_P(n^{-1/2}) \\ &= O(h) + O_P(n^{-1/2}) = o_P((n\alpha h^2)^{-1/2}). \end{aligned}$$

When $\alpha = \alpha_n$ is an intermediate sequence, a tail analogue of the above statement can be derived by using, e.g., Theorem 2.4.8 in [10] for univariate regularly varying distributions in such a way that

$$\begin{aligned} \frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} &= (1-h)^{-1/\gamma_1} - (1+h)^{-1/\gamma_1} + o_P((n\alpha)^{-1/2}) + o(A(1/\alpha)) \\ &= O(h) + o_P((n\alpha)^{-1/2}) + o(A(1/\alpha)) = o_P((n\alpha h^2)^{-1/2}) \end{aligned}$$

as $n \to \infty$. $\square$

Lemma 3.6.4. *Under the conditions of Theorem 3.2.1 or Theorem 3.2.2, as $n \to \infty$,*

$$\sqrt{n\alpha}\left(\frac{\widehat{C}(\alpha,\alpha)-\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)}, \frac{\widehat{ES}_\alpha(X)-\widetilde{ES}_\alpha(X)}{ES_\alpha(X)}, \frac{\widehat{ES}_\alpha(Y)-\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)}\right) \xrightarrow{P} (0,0,0). \tag{3.11}$$

Proof. The convergence in the first coordinate for fixed α is already noticed in [8], with Lemma 3.6.1 and applications of the delta method; see also, e.g., Lemma 1 in [59] for details. The proofs for intermediate α is very much the same by using Lemma 3.6.2 instead and we refer to Lemma 1 in [58] for more details.

In the following we only prove the convergence in the second coordinate since the proof for the third coordinate is completely analogous. Noting $0 < Q_1(1-\alpha) \le ES_\alpha(X)$, it suffices to show

$$\sqrt{n\alpha}\left(\frac{\widehat{ES}_\alpha(X)-\widetilde{ES}_\alpha(X)}{Q_1(1-\alpha)}\right) \xrightarrow{P} 0. \tag{3.12}$$

Write

$$\begin{aligned}\frac{\widehat{ES}_\alpha(X)-\widetilde{ES}_\alpha(X)}{Q_1(1-\alpha)} =& \frac{1}{n\alpha}\sum_{i=1}^n \frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) - \mathbb{1}[\bar{F}_{n1}(X_i)<\alpha]\right)\\ &+ \frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(1-\frac{1}{n\alpha}\sum_{i=1}^n \mathbb{1}[\bar{F}_{n1}(X_i))<\alpha]\right) =: J_1 + J_2.\end{aligned}$$

Applying Lemma 3.6.3 yields

$$\begin{aligned}|J_1| &\le \frac{X_{n-[n\alpha(1-h)]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)}\frac{1}{n\alpha}\sum_{i=1}^n\left|K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) - \mathbb{1}[\bar{F}_{n1}(X_i)<\alpha]\right|\\ &\le \frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)}\cdot\frac{1}{n\alpha}\cdot O_P(n\alpha h) = o_P((n\alpha)^{-1/2}).\end{aligned}$$

The rest is straightforward since

$$|J_2| = \frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(1-\frac{\lceil n\alpha\rceil - 1}{n\alpha}\right) = O_P((n\alpha)^{-1}) = o_P((n\alpha)^{-1/2}).$$

$\square$

Proof of Theorem 3.2.1. Write

$$\begin{cases} \frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1 = \left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1\right) \frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} \frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)} + \left(\frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1\right) \frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)} - \frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)} \left(\frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right), \\ \frac{\widehat{\rho}_\alpha}{\rho_\alpha} - 1 = \left(\frac{\widehat{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1\right) \frac{\widehat{ES}_\alpha(X)}{ES_\alpha(X)} \frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)} + \left(\frac{\widehat{ES}_\alpha(X)}{ES_\alpha(X)} - 1\right) \frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)} - \frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)} \left(\frac{\widehat{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right). \end{cases} \tag{3.13}$$

Combining this with Lemma 3.6.4 yields that

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha}{\rho_\alpha} - \frac{\widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{P} 0.$$

For the rest it remains to show that $\sqrt{n\alpha}(\widetilde{\rho}_\alpha/\rho_\alpha - 1) \xrightarrow{d} N(0, \sigma_\alpha^2)$, or, using (3.13), to show that

$$\sqrt{n\alpha}\left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1, \frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1, \frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right) \xrightarrow{d} (\Lambda_\alpha, \Theta_{\alpha,1}, \Theta_{\alpha,2}). \tag{3.14}$$

For convenient presentation, all the limit processes below are defined on the same probability space, via the Skorohod construction. However, they are only equal in distribution to the original processes. Using Lemma 3.6.1 with some $\eta \in \left(\frac{1}{2+\delta}, \frac{1}{2}\right)$ we have

$$\left\{\frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x \wedge \alpha y)},\ x, y \in [0, 1/\alpha]\right\} \xrightarrow{a.s.} \left\{\frac{1}{\sqrt{\alpha}} \frac{B_C(\alpha x, \alpha y)}{q_\eta(\alpha x \wedge \alpha y)},\ x, y \in [0, 1/\alpha]\right\}. \tag{3.15}$$

Applying the inverse lemma from [77] (or see Lemma A.0.2 in [10]) on the marginal processes $\mathbb{W}_n(\cdot, 1/\alpha)$ and $\mathbb{W}_n(1/\alpha, \cdot)$ around a neighborhood of 1 yields that

$$\begin{aligned} \sqrt{n\alpha}(e_n - 1, e_n' - 1) \quad &:= \quad \sqrt{n\alpha}\left(\check{R}_{n1}^{\leftarrow}(1) - 1, \check{R}_{n2}^{\leftarrow}(1) - 1\right) \\ &\xrightarrow{a.s.} \quad \left(-\tfrac{1}{\sqrt{\alpha}} B_C(\alpha, 1), -\tfrac{1}{\sqrt{\alpha}} B_C(1, \alpha)\right), \end{aligned} \tag{3.16}$$

where $\check{R}_{n1}(\cdot) = R_n(\cdot, 1/\alpha) = \frac{1}{\alpha}\check{C}(\alpha\cdot, 1)$ and $\check{R}_{n2}(\cdot) = R_n(1/\alpha, \cdot) = \frac{1}{\alpha}\check{C}(1, \alpha\cdot)$.

Using (3.15) and (3.16) we then have, cf. pages 52-53 in [62],

$$
\begin{aligned}
&\sqrt{n\alpha}\left(\frac{\check{C}_n(\alpha,\alpha)}{C(\alpha,\alpha)}-1\right)\\
&\overset{a.s.}{=}\frac{\alpha}{C(\alpha,\alpha)}\left\{\mathbb{W}_n(1,1)+C_1(\alpha,\alpha)\sqrt{n\alpha}\,(e_n-1)+C_2(\alpha,\alpha)\sqrt{n\alpha}\,(e_n'-1)\right\}+o(1)\\
&\xrightarrow{a.s.}\Lambda_\alpha.
\end{aligned}
\tag{3.17}
$$

Moreover, since we can write

$$
\begin{cases}
\widetilde{ES}_\alpha(X)\overset{a.s.}{=}-\int_0^{e_n}\check{R}_n(u,\infty)dQ_1(1-\alpha u)+Q_1(1-e_n\alpha),\\
ES_\alpha(X)=-\int_0^1 udQ_1(1-\alpha u)+Q_1(1-\alpha),
\end{cases}
$$

and, therefore,

$$
\begin{aligned}
&\sqrt{n\alpha}\left(\widetilde{ES}_\alpha(X)-ES_\alpha(X)\right)-ES_\alpha(X)\Theta_{\alpha,1}\\
&\overset{a.s.}{=}-\int_0^1\left(\mathbb{W}_n(x,1/\alpha)-\frac{1}{\sqrt{\alpha}}B_C(\alpha x,1)\right)dQ_1(1-\alpha x)\\
&\quad+\int_{e_n}^1\left(\sqrt{n\alpha}(1-x)+\mathbb{W}_n(x,\infty)\right)dQ_1(1-\alpha x)\\
&\le\sup_{x\in(0,1)}\frac{\left|\mathbb{W}_n(x,1/\alpha)-\frac{1}{\sqrt{\alpha}}B_C(\alpha x,1)\right|}{q_\eta(\alpha x)}\int_0^1(\alpha x)^\eta dQ_1(1-\alpha x)\\
&\quad+\sup_{|x-1|\le|e_n-1|}\left|\sqrt{n\alpha}(1-x)+\mathbb{W}_n(x,\infty)\right|\cdot|Q_1(1-\alpha e_n)-Q_1(1-\alpha)|\\
&\xrightarrow{a.s.}0+0=0.
\end{aligned}
$$

Similarly, we also have

$$
\sqrt{n\alpha}\left(\widetilde{ES}_\alpha(Y)-ES_\alpha(Y)\right)-ES_\alpha(Y)\Theta_{\alpha,2}\xrightarrow{a.s.}0.
$$

Statement (3.14) then follows. $\square$

Proof of Theorem 3.2.2. The proof is analogous to that of Theorem 3.2.1, by replacing $\mathbb{W}_n(x, 1/\alpha)$ by $\mathbb{W}_n(x, \infty)$, $\mathbb{W}_n(1/\alpha, y)$ by $\mathbb{W}_n(\infty, y)$ (since $\bar{F}_1(X_i) < 1$ and $\bar{F}_2(Y_i) < 1$ for all $i = 1, \ldots, n$), and the processes $\frac{1}{\sqrt{\alpha}} B_C(\alpha x, \alpha y)$ by $W_R(x, y)$, $\frac{1}{\sqrt{\alpha}} B_C(\alpha x, 1)$ by $W_R(x, \infty)$, and $\frac{1}{\sqrt{\alpha}} B_C(1, \alpha y)$ by $W_R(\infty, y)$. Particularly, with Lemma 3.6.2 we can show that

$$\sqrt{n\alpha}\left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1, \frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1, \frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right) \xrightarrow{d} (\Lambda_0, \Theta_{0,1}, \Theta_{0,2}). \tag{3.18}$$

While the proof of the first coordinate-wise convergence is straightforward by recalling the equality in (3.17), that of the second and third ones are provided in details in the proof of Proposition 3 in [63] (as a special case by taking $X = Y$ therein) and thus omitted. The rest follows from (3.13) and Lemma 3.6.4, as in the proof of Theorem 3.2.1. $\square$

To show Theorem 3.2.3, we need some intermediate lemmas for the component-wise jackknife pseudo samples as, for $i = 1, \ldots, n$,

$$\begin{cases} \widehat{V}_{C,i} = n\widehat{C}(\alpha,\alpha) - (n-1)\widehat{C}_i(\alpha,\alpha), \\ \widehat{V}_{X,i} = n\widehat{ES}_\alpha(X) - (n-1)\widehat{ES}_{\alpha,i}(X), \\ \widehat{V}_{Y,i} = n\widehat{ES}_\alpha(Y) - (n-1)\widehat{ES}_{\alpha,i}(Y). \end{cases}$$

Specifically, we shall first develop the joint asymptotics of the jackknife means and jackknife (co)variance based on these component-wise pseudo samples. The (marginal) results below for $\widehat{V}_{C,1}, \ldots, \widehat{V}_{C,n}$ are taken mostly from [58] for intermediate α and [59] for fixed α.

Lemma 3.6.5. *As $n \to \infty$, under the conditions of Theorem 3.2.1 or Theorem 3.2.2,*

$$\sqrt{n\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)} - 1, \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)} - 1, \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)} - 1\right) \xrightarrow{P} (0,0,0). \tag{3.19}$$

Proof. The convergence in the first coordinate is a direct consequence of Lemma 2 in [59] and Lemma 2 in [58] under the conditions of, respectively, Theorem 3.2.1 and Theorem 3.2.2 by noting that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}-1=\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\alpha}\widehat{V}_{C,i}-\frac{1}{\alpha}\widehat{C}(\alpha,\alpha)\right\}\left(\frac{C(\alpha,\alpha)}{\alpha}\right)^{-1}\frac{C(\alpha,\alpha)}{\widehat{C}(\alpha,\alpha)}=o_P((n\alpha)^{-1/2}).$$

In the following we shall only prove the convergence in the second coordinate since the proof for the third coordinate is completely analogous. A crucial step is to observe that

$$\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)=\frac{1}{n-1}\left(\bar{F}_{n1}(x)-\mathbb{1}[x<X_i]\right)\quad\text{for}\quad x\in\mathbb{R},\tag{3.20}$$

which implies that

$$\sum_{i=1}^{n}\left(\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)\right)=0\quad\text{for}\quad x\in\mathbb{R},\tag{3.21}$$

$$\max_{1\le i\le n}\sup_{x\in\mathbb{R}}\left|\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)\right|\le n^{-1}.\tag{3.22}$$

Now write

$$\begin{aligned}\widehat{V}_{X,i}&=n\widehat{ES}_{\alpha}(X)-(n-1)\widehat{ES}_{\alpha,i}(X)\\&=\frac{1}{\alpha}\sum_{j=1}^{n}\left(X_j-X_{n-\lceil n\alpha\rceil:n}\right)\left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)-K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right)\\&+\left\{\frac{1}{\alpha}\left(X_i-X_{n-\lceil n\alpha\rceil:n}\right)K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)+X_{n-\lceil n\alpha\rceil:n}\right\}\\&=:\widehat{V}_{X,i,1}+\widehat{V}_{X,i,2}.\end{aligned}\tag{3.23}$$

Applying the mean value theorem yields that, for each pair (i,j), there exists $\varepsilon_{i,j}$ between

$\bar{F}_{n1}(X_j)$ and $\bar{F}_{n1,i}(X_j)$ such that

$$\begin{aligned}
&K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) \\
&= k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h} \\
&+ \frac{1}{2}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right)^2 .
\end{aligned} \tag{3.24}$$

Note that (3.22) implies that

$$\sup_{1\le i\le n} \mathbb{1}\left(\left|\frac{1-\varepsilon_{i,j}/\alpha}{h}\right| \le 1\right) \le \mathbb{1}\left(1-h-\frac{1}{n\alpha} \le \frac{\bar{F}_{n1}(X_j)}{\alpha} \le 1+h+\frac{1}{n\alpha}\right). \tag{3.25}$$

It follows from (3.20) and (3.21) that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,1} =& \frac{1}{n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right) k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\left(\sum_{i=1}^{n}\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right) \\
&+ \frac{1}{2n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n} k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right)^2 \\
=& \frac{1}{2n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n} k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)-\mathbb{1}[X_j<X_i]}{(n-1)\alpha h}\right)^2 \\
=& \frac{1}{2n(n-1)^2\alpha h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n} k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)}{\alpha}\right)^2 \\
&- \frac{1}{n(n-1)^2\alpha h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n} k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)\mathbb{1}[X_j<X_i]}{\alpha^2}\right) \\
&+ \frac{1}{2n(n-1)^2\alpha^2 h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n} k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\frac{\mathbb{1}[X_j<X_i]}{\alpha} \\
=&: J_1 - J_2 + J_3.
\end{aligned}$$

We start with the most difficult term J_3. We have that, for some $M > 0$

$$
\begin{aligned}
&\left|\frac{J_3}{Q_1(1-\alpha)}\right| \\
&\leq \frac{M}{n^2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right)\frac{1}{n\alpha}\sum_{i=1}^{n}\mathbb{1}[X_j < X_i] \\
&\leq \frac{M}{n^2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right)\left(\frac{\bar{F}_{n1}(X_j)}{\alpha}\right) \\
&\leq \frac{M}{n^2\alpha^2h^2}\cdot(2n\alpha h+3)\frac{X_{n+1-[n\alpha(1+h)]:n} - X_{n-1-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)}\left(1+h+\frac{1}{n\alpha}\right) \\
&= O((n\alpha h)^{-1})\cdot o_P((n\alpha h^2)^{-1/2}) = o_P((n\alpha)^{-1/2}).
\end{aligned}
$$

Similarly we can show that

$$\left|\frac{J_1}{Q_1(1-\alpha)}\right| = o_P((n\alpha)^{-1/2}), \text{ and } \left|\frac{J_2}{Q_1(1-\alpha)}\right| = o_P((n\alpha)^{-1/2}).$$

Hence,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right| \leq \left|\frac{J_1}{Q_1(1-\alpha)}\right| + \left|\frac{J_2}{Q_1(1-\alpha)}\right| + \left|\frac{J_3}{Q_1(1-\alpha)}\right| = o_P((n\alpha)^{-1/2}). \tag{3.26}$$

Now, using the mean value theorem (again), we know there exists an $\widetilde{\varepsilon}_{i,j}$ between $\bar{F}_{n1,i}(X_j)$ and $\bar{F}_{n1}(X_j)$ such that

$$K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) = k\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right)\frac{\bar{F}_{n1}(X_j)-\bar{F}_{n1,i}(X_j)}{\alpha h} \tag{3.27}$$

and, similar as (3.25),

$$\sup_{1\leq i\leq n}\mathbb{1}\left(\left|\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right| \leq 1\right) \leq \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right). \tag{3.28}$$

It follows that, for some $M > 0$,

$$
\begin{aligned}
&\frac{1}{Q_1(1-\alpha)}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,2}-\widehat{ES}_\alpha(X)\right)\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)-K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)\right)\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}k\left(\frac{1-\widetilde{\varepsilon}_{i,i}/\alpha}{h}\right)\frac{\bar{F}_{n1,i}(X_i)-\bar{F}_{n1}(X_i)}{\alpha h}\\
=&\frac{1}{n(n-1)\alpha h}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}k\left(\frac{1-\widetilde{\varepsilon}_{i,i}/\alpha}{h}\right)\frac{\bar{F}_{n1}(X_i)}{\alpha}\\
\leq&\frac{M}{n^2\alpha h}\sum_{i=1}^{n}\frac{\left|X_i-X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n\alpha}\leq\frac{\bar{F}_{n1}(X_i)}{\alpha}\leq 1+h+\frac{1}{n\alpha}\right)\frac{\bar{F}_{n1}(X_i)}{\alpha}\\
\leq&\frac{M}{n^2\alpha h}(2n\alpha h+3)\frac{X_{n+1-[n\alpha(1+h)]:n}-X_{n-1-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)}\left(1+h+\frac{1}{n\alpha}\right)\\
=&o_P(n^{-1})\cdot o_P((n\alpha h^2)^{-1/2})\cdot O(1)=o_P((n\alpha)^{-1/2}).
\end{aligned}
$$

Therefore

$$
\frac{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i}-\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}=\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}+\frac{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,2}-\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}=o_P((n\alpha)^{-1/2}).
$$

The rest follows from the simple fact that $\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)}=\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\cdot\frac{ES_\alpha(X)}{\widehat{ES}_\alpha(X)}=O_P(1)$. ☐

Lemma 3.6.6. *As $n\to\infty$, under the conditions of Theorem 3.2.3*

$$
\max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|=O_P(1),\quad \max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|=o_P((n\alpha)^{1/2}),
$$

and

$$
\max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|=o_P((n\alpha)^{1/2}).
$$

Proof. From the proofs of Theorem 2 in [58] and Theorem 2 in [59], we have $\max_{1\leq i\leq n}|\widehat{V}_{C,i}|=O_p(1)$ for both intermediate and fixed α. The first claim then follows from the consistency of $\widehat{C}(\alpha,\alpha)/\alpha$ (implied by Theorems 3.2.1 and 3.2.2 above).

Next we shall show the second claim. Recall from (3.23) that we can write $\widehat{V}_{X,i} = \widehat{V}_{X,i,1} + \widehat{V}_{X,i,2}$. With (3.22) and (3.25), we have that, using the Taylor expansion (3.24), for some large $M > 0$

$$\begin{aligned}
&\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right| \\
\le&\frac{1}{\alpha h}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\max_{1\le i\le n}\left|\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)\right| \\
&+\frac{1}{2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right|\max_{1\le i\le n}\left|k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\right|\max_{1\le i\le n}\left(\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)\right)^2 \\
\le&\frac{M}{n\alpha h}\sum_{j=1}^{n}\frac{\left|X_j - X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n}\le\frac{\bar{F}_{n1}(X_j)}{\alpha}\le 1+h+\frac{1}{n}\right) \\
&+\frac{M}{2n^2\alpha^2h^2}\sum_{j=1}^{n}\frac{\left|X_j - X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n}\le\frac{\bar{F}_{n1}(X_j)}{\alpha}\le 1+h+\frac{1}{n}\right) \\
\le&\left\{\frac{M}{n\alpha h}+\frac{M}{2n^2\alpha^2h^2}\right\}\cdot(2n\alpha h+3)\frac{X_{n-[n\alpha(1+h)]:n}-X_{n-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)} \\
=&O(1)\cdot O_P((n\alpha h^2)^{-1/2}) = o_P(1),
\end{aligned}$$

where for the last line we apply Lemma 3.6.3. It remains to verify that

$$\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\right| = o_P((n\alpha)^{1/2}). \tag{3.29}$$

Recall from Lemma 11.2 in [1] that

$$\max_{1\le i\le n}|X_i| = o_P(n^{1/2}).$$

When $\alpha\in(0,1)$ is fixed, $\widehat{ES}_\alpha(X)\xrightarrow{P} ES_\alpha(X) > 0$ and Lemma 11.2 in [1] yields that

$$\max_{1\le i\le n}\left|\alpha\widehat{V}_{X,i,2}\right| \le \max_{1\le i\le n}|X_i| + \left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha) = o_P(n^{1/2}) + O_P(1) = o_P((n\alpha)^{1/2}).$$

When α is an intermediate sequence, similarly, we have

$$\begin{aligned}\max_{1\le i\le n}\left|\alpha\widehat{V}_{X,i,2}\right| &\le \max_{1\le i\le n}|X_i|\mathbb{1}[X_i > X_{n-\lceil n\alpha(1+h)\rceil:n}] + \left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha)\\ &\le \max\{|X_{n:n}|, \left|X_{n-\lceil n\alpha(1+h)\rceil:n}\right|\} + \left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha)\,.\end{aligned}$$

A fundamental result in extreme value theory (see, e.g., Section 1.1 in [10]) tells that

$$X_{n:n} = O_p(Q_1(1-1/n)).$$

Therefore, in conjunction with the regular variation of Q_1 implied by Assumption (2.a),

$$\begin{aligned}\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right| \le& \frac{\max\{|X_{n:n}|, \left|X_{n-\lceil n\alpha(1+h)\rceil:n}\right|\}}{Q_1(1-\alpha)} + (\alpha+1)\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\\ =& O_P\left(\frac{Q_1(1-1/n)}{Q_1(1-\alpha)}\right) + O_P(1) = o_P((n\alpha)^{\gamma_1}) = o_P((n\alpha)^{1/2})\end{aligned}$$

since $\gamma_1 \in (0, 1/2)$. Recalling $\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)} = \frac{Q_1(1-\alpha)}{ES_\alpha(X)} \cdot \frac{ES_\alpha(X)}{\widehat{ES}_\alpha(X)} = O_P(1)$, the second part of the lemma follows. The proof for the third part is completely analogous and hence omitted. □

Lemma 3.6.7. *Under Assumptions 3.2.2 and 3.2.3, as $\alpha \downarrow 0$,*

$$\Sigma^{(\alpha)} := \textit{Cov}(\Lambda_\alpha, \Theta_{\alpha,1}, \Theta_{\alpha,2}) \to \textit{Cov}\left(\Lambda_0, \Theta_{0,1}, \Theta_{0,2}\right).$$

Proof. It is easy to verify that $\mathrm{Var}(\Lambda_\alpha) \to \mathrm{Var}(\Lambda_0)$. Moreover,

$$\begin{aligned}\mathrm{Cov}(\Theta_{\alpha,1}, \Theta_{\alpha,2}) = \left(\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)^2 \Bigg\{ &\int_0^1\int_0^1 \left(\frac{1}{\alpha}C(\alpha x, \alpha y)\right) d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right) d\left(\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}\right)\\ &-\alpha\int_0^1\int_0^1 xy d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right) d\left(\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}\right)\Bigg\}.\end{aligned}$$

Note that $\sup_{0<x,y\le 1}\frac{\left|\alpha^{-1}C(\alpha x,\alpha y) - R(x,y)\right|}{(x\wedge y)^\tau} \to 0$ by assumptions, and by Potter's inequality in

[78] (or see Proposition B.1.9 in [10]) we have that

$$\sup_{\mathbf{x}\in(0,1)} x^{\tau}\left|\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}-x^{-\gamma_1}\right|\to 0, \text{ and } \sup_{\mathbf{y}\in(0,1)} y^{\tau}\left|\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}-y^{-\gamma_1}\right|\to 0. \quad (3.30)$$

By some calculations we can therefore show that

$$\text{Cov}(\Theta_{\alpha,1},\Theta_{\alpha,2})\to(1-\gamma_1)^2\left(\int_0^1\int_0^1 R(x,y)dx^{-\gamma_1}dy^{-\gamma_2}-0\right)=\text{Cov}(\Theta_{0,1},\Theta_{0,2}),$$

where we also apply the Karamata's theorem (see, e.g., Theorem B.1.5 in [10]) which says that

$$\frac{Q_1(1-\alpha)}{ES_\alpha(X)}=\frac{Q_1(1-\alpha)}{\int_0^1 Q_1(1-\alpha u)du}\to-\gamma_1+1 \quad (3.31)$$

Similarly, even easier, we can show $\text{Var}(\Theta_{\alpha,j})\to\text{Var}(\Theta_{0,j})$ for $j=1,2$, and

$$\begin{aligned}
&\text{Cov}(\Lambda_\alpha,\Theta_{\alpha,1})\\
=&-\frac{\alpha}{C(\alpha,\alpha)}\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\Bigg\{\int_0^1\left(\frac{C(\alpha x,\alpha)}{\alpha}-C_1(\alpha,\alpha)x-C_2(\alpha,\alpha)\frac{C(\alpha x,\alpha)}{\alpha}\right)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)\\
&-(C(\alpha,\alpha)-\alpha C_1(\alpha,\alpha)-\alpha C_2(\alpha,\alpha))\int_0^1 xd\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)\Bigg\}\\
\to&\frac{\gamma_1-1}{R(1,1)}\int_0^1\left(R(x,1)-R_1(1,1)x-R_2(1,1)R(x,1)\right)dx^{-\gamma_1}=\text{Cov}(\Lambda_0,\Theta_{0,1}).
\end{aligned}$$

The proof of $\text{Cov}(\Lambda_\alpha,\Theta_{\alpha,2})\to\text{Cov}(\Lambda_0,\Theta_{0,2})$ is completely analogous. □

The following lemma establishes the consistency of the jackknife covariance matrix of the relative estimation errors of component-wise nonparametric estimators, where the smoothing technique plays an important role.

Lemma 3.6.8. *Denote* $\widehat{\mathbf{V}}_i=\left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)},\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)},\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)$, $i=1,\ldots,n$. *Under the conditions* *of Theorem 3.2.1 or Theorem 3.2.2, as $n \to \infty$,*

$$\widehat{\Sigma} - \Sigma^{(\alpha)} := \frac{\alpha}{n} \sum_{i=1}^{n} \left(\widehat{\mathbf{V}}_i - \mathbf{1} \right) \left(\widehat{\mathbf{V}}_i - \mathbf{1} \right)' - \Sigma^{(\alpha)} \xrightarrow{P} 0.$$

Proof. We only prove the convergence of the (co)variance terms $\widehat{\Sigma}_{1,2}$, $\widehat{\Sigma}_{1,3}$, $\widehat{\Sigma}_{2,3}$, $\widehat{\Sigma}_{2,2}$, and $\widehat{\Sigma}_{3,3}$. The convergence of $\widehat{\Sigma}_{2,1}$, $\widehat{\Sigma}_{3,1}$ and $\widehat{\Sigma}_{3,2}$ then follows by the symmetry of $\widehat{\Sigma}$ (and Σ), and the convergence of $\widehat{\Sigma}_{1,1}$ is readily known by Lemma 3 in [58] for intermediate α and Lemma 3 in [59] for fixed α.

Consistency of $\widehat{\Sigma}_{1,2}$ and $\widehat{\Sigma}_{1,3}$. Recall from (3.23) that $\widehat{V}_{X,i} = \widehat{V}_{X,i,1} + \widehat{V}_{X,i,2}$. Using the Taylor expansion (3.27), with (3.22), (3.28) and Lemma 3.6.3, we have that for some large $M > 0$

$$\begin{aligned}
&\frac{\alpha}{n} \sum_{i=1}^{n} \frac{\widehat{V}_{X,i,1}^2}{Q_1^2(1-\alpha)} \\
&= \frac{\alpha}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \right) \left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h} \right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h} \right) \right) \right)^2 \\
&\le \alpha \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \right)^2 \left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h} \right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h} \right) \right)^2 \\
&= \frac{1}{\alpha h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \right)^2 k^2\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h} \right) \left(\bar{F}_{n1,i}(X_j) - \bar{F}_{n1}(X_j) \right)^2 \\
&\le \frac{M}{n\alpha h^2} \sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \right)^2 \mathbb{1}\left(1 - h - \frac{1}{n} \le \frac{\bar{F}_{n1}(X_j)}{\alpha} \le 1 + h + \frac{1}{n} \right) \\
&\le \frac{M}{n\alpha h^2} (2n\alpha h + 3) \cdot \left(\frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} \right)^2 = o_P((n\alpha)^{-1}) = o_P(1).
\end{aligned}$$

Write $\widehat{V}_{C,i} = \widehat{V}_{C,i,1} + \widehat{V}_{C,i,2}$, where

$$\begin{cases} \widehat{V}_{C,i,1} = \sum_{j=1}^{n} \left\{ K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h} \right) K\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h} \right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h} \right) K\left(\frac{1-\bar{F}_{n2,i}(Y_j)/\alpha}{h} \right) \right\}, \\ \widehat{V}_{C,i,2} = K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h} \right) K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h} \right). \end{cases}$$

We have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,2}\cdot\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\\
=&\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)K^2\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h}\right)\\
&+\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h}\right)\\
=&\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\mathbb{1}\left(X_i>X_{n-\lceil n\alpha\rceil:n},Y_i>Y_{n-\lceil n\alpha\rceil:n}\right)+\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\widehat{C}(\alpha,\alpha)+o_P(1)\\
=&\int_{e_n}^{0}\check{R}_n(u,e_n')d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)+C(\alpha,\alpha)+o_P(1)\\
=&-\int_{0}^{1}R_n(u,1)d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)+C(\alpha,\alpha)+o_P(1),
\end{aligned}
$$

where in the last step we recall from the proofs of Theorems 3.2.1 and 3.2.2 that $e_n := R_{n1}^{\leftarrow}(1)=1+o_P(1)$ and $e_n' := R_{n2}^{\leftarrow}(1)=1+o_P(1)$.

Moreover, similar to the proof of (3.26), we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,1}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)+\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right\}\\
&\times\frac{\bar{F}_{n2,i}(Y_j)-\bar{F}_{n2}(Y_j)}{\alpha h}K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)\\
&+\frac{1}{n\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)+\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right\}\\
&\times\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)+o_P(1)\\
=:&\,T_1+T_2+o_P(1).
\end{aligned}
$$

Note that $\bar{F}_{n2,i}(y)-\bar{F}_{n2}(y)=\frac{1}{n-1}\left(\bar{F}_{n2}(y)-\mathbb{1}\left[\bar{F}_{n2}(Y_i)<\bar{F}_{n2}(y)\right]\right)$. Therefore, in conjunction with (3.14), (3.18) and (3.31), we have

$$
\begin{aligned}
T_1 =& \frac{\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)} \frac{1}{(n-1)\alpha h} \sum_{j=1}^{n} \bar{F}_{n2}(Y_j) k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) \\
& - \frac{1}{(n-1)\alpha h} \sum_{j=1}^{n} \left\{ \frac{1}{n\alpha} \sum_{i=1}^{n} \frac{X_i - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) \mathbb{1}\left[\bar{F}_{n2}(Y_i) < \bar{F}_{n2}(Y_j)\right] \right\} \\
& \quad \times k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) \\
& - \alpha \frac{X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \frac{1}{(n-1)\alpha h} \sum_{j=1}^{n} \left\{ \frac{1}{n\alpha} \sum_{i=1}^{n} \mathbb{1}\left[\bar{F}_{n2}(Y_i) < \bar{F}_{n2}(Y_j)\right] \right\} \times \\
& \quad k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) \\
& = C_2(\alpha,\alpha) \frac{ES_\alpha(X)}{Q_1(1-\alpha)} \alpha + C_2(\alpha,\alpha) \int_0^1 R_n(u,1) d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right) - \alpha C_2(\alpha,\alpha) + o_P(1) \\
& = \alpha C_2(\alpha,\alpha) \left\{ \frac{ES_\alpha(X)}{Q_1(1-\alpha)} - 1 \right\} + C_2(\alpha,\alpha) \int_0^1 R_n(u,1) d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right) + o_P(1)
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
T_2 &= C_1(\alpha,\alpha) \left\{ \alpha \frac{ES_\alpha(X)}{Q_1(1-\alpha)} - \frac{ES_\alpha(X)}{Q_1(1-\alpha)} + 1 - \alpha \right\} + o_P(1) \\
&= -C_1(\alpha,\alpha)(1-\alpha) \left\{ \frac{ES_\alpha(X)}{Q_1(1-\alpha)} - 1 \right\} + o_P(1).
\end{aligned}
$$

To conclude we have, again recalling (3.14), (3.18) and (3.31),

$$
\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\\
=&\frac{\alpha}{\widehat{C}(\alpha,\alpha)}\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)}\left\{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}+\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,1}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right.\\
&\left.+\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,2}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right\}\\
=&-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\int_0^1\frac{R_n(u,1)}{R_n(1,1)}d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)-C_1(\alpha,\alpha)\frac{1}{R_n(1,1)}\left\{1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right\}\\
&+C_2(\alpha,\alpha)\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\int_0^1\frac{R_n(u,1)}{R_n(1,1)}d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)\\
&-\alpha\left(1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)\left\{1-\frac{C_1(\alpha,\alpha)}{R_n(1,1)}-\frac{C_2(\alpha,\alpha)}{R_n(1,1)}\right\}+\alpha+o_P(1)\\
=&\Sigma_{1,2}^{(\alpha)}+\alpha+o_P(1).
\end{aligned}
$$

It follows that, in conjunction with Lemma 3.6.5, as $n\to\infty$

$$
\begin{aligned}
\widehat{\Sigma}_{1,2}=&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\alpha\\
=&\Sigma_{1,2}^{(\alpha)}+\alpha-\alpha-\alpha+\alpha+o_P(1)=\Sigma_{1,2}^{(\alpha)}+o_P(1).
\end{aligned}
$$

Similarly, we can show that $\widehat{\Sigma}_{1,3}=\Sigma_{1,3}^{(\alpha)}+o_P(1)$.

Consistency of $\widehat{\Sigma}_{2,3}$, $\widehat{\Sigma}_{2,2}$ and $\widehat{\Sigma}_{2,2}$. Next, we write $\widehat{V}_{X,i}=\widehat{V}_{X,i,1}+\widehat{V}_{X,i,2}$ as defined in equation (3.23). Analogously we define $\widehat{V}_{Y,i,1}$ and $\widehat{V}_{Y,i,2}$. With (3.27) and (3.28), there exists some $M > 0$ such that for all $i = 1, \ldots, n$

$$\begin{aligned}
\left|\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right| \leq & \sum_{j=1}^{n} \left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \left|K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right| \\
= & \sum_{j=1}^{n} \left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| k\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right) \frac{\left|\bar{F}_{n1,i}(X_j) - \bar{F}_{n1}(X_j)\right|}{\alpha h} \\
\leq & \frac{M}{n\alpha h} \sum_{j=1}^{n} \left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \mathbb{1}\left(1 - h - \frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1 + h + \frac{1}{n\alpha}\right) \\
\leq & \frac{M}{n\alpha h}(2n\alpha h + 3) \frac{X_{n-[n\alpha(1-\alpha/h)]:n} - X_{n-[n\alpha(1+\alpha/h)]:n}}{Q_1(1-\alpha)}.
\end{aligned}$$

It follows that, together with Lemma 3.6.3,

$$\begin{aligned}
\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{X,i,1}}{ES_\alpha(X)}\right)^2 \leq & \frac{\alpha M^2}{n^2\alpha^2h^2}(2n\alpha h + 3)^2 \left(\frac{X_{n-[n\alpha(1-\alpha/h)]:n} - X_{n-[n\alpha(1+\alpha/h)]:n}}{Q_1(1-\alpha)}\right)^2 \\
= & O(\alpha) \cdot o_P((n\alpha h^2)^{-1/2}) = o_P(1).
\end{aligned}$$

Similarly, we also have $\frac{\alpha}{n}\sum_{i=1}^n\left(\frac{\widehat{V}_{Y,i,1}}{ES_\alpha(X)}\right)^2=o_P(1)$. Moreover, we have

$$
\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^n\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\frac{\widehat{V}_{Y,i,2}}{Q_2(1-\alpha)}\\
=&\frac{1}{n\alpha}\sum_{i=1}^n\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\left(\frac{Y_i-Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\right)K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right)\\
&+\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{\widehat{ES}_\alpha(Y)}{Q_2(1-\alpha)}+\alpha\frac{Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\frac{\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}-\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\\
=&\frac{1}{n\alpha}\sum_{i=1}^n\left(\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\left(\frac{Y_i-Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\right)\mathbb{1}\left[X_j>X_{n-\lceil n\alpha\rceil:n},Y_j>Y_{n-\lceil n\alpha\rceil:n}\right]\\
&+\alpha\frac{ES_\alpha(Y)}{Q_2(1-\alpha)}+\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_P(1)\\
=&\int_0^{e_n}\int_0^{e_n'}\check{R}_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)\\
&+\alpha\frac{ES_\alpha(Y)}{Q_2(1-\alpha)}+\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_P(1)\\
=&\int_0^1\int_0^1 R_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)\\
&+\alpha\frac{ES_\alpha(Y)}{Q_2(1-\alpha)}+\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_P(1),
\end{aligned}
$$

where in the last step we apply Lemma 3.6.1 for fixed α with $\eta\in\left(\frac{1}{2+\delta},\frac{1}{2}\right)$ and Lemma 3.6.2 for intermediate α with $\eta\in(\gamma_1,1/2)$.

It follows that, again recalling (3.14), (3.18) and (3.31),

$$
\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^n\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}\\
=&\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\frac{Q_2(1-\alpha)}{ES_\alpha(Y)}\int_0^1\int_0^1 R_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)\\
&-\alpha\left(1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)\left(1-\frac{Q_2(1-\alpha)}{ES_\alpha(Y)}\right)+\alpha+o_P(1)\\
=&\Sigma_{2,3}^{(\alpha)}+\alpha+o_P(1).
\end{aligned}
$$

Analogously, we can also show that

$$\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\right)^2=\Sigma_{2,2}^{(\alpha)}+\alpha+o_P(1),\ \text{and}\ \frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}\right)^2=\Sigma_{3,3}^{(\alpha)}+\alpha+o_P(1).$$

Hence, in conjunction with Lemma 3.6.5,

$$\begin{aligned}\widehat{\Sigma}_{2,3}=&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}+\alpha+o_P(1)\\ =&\Sigma_{2,3}^{(\alpha)}+\alpha-\alpha-\alpha+\alpha+o_P(1)=\Sigma_{2,3}^{(\alpha)}+o_P(1).\end{aligned}$$

Similarly, $\widehat{\Sigma}_{2,2}=\Sigma_{2,2}^{(\alpha)}+o_P(1)$ and $\widehat{\Sigma}_{3,3}=\Sigma_{3,3}^{(\alpha)}+o_P(1)$. □

Finally, we combine the above component-wise results to establish the asymptotics of the jackknife pseudo sample of our relative risk measure, that is, $\widehat{V}_{\rho,1},\ldots,\widehat{V}_{\rho,n}$.

Lemma 3.6.9. *As $n\to\infty$, under the conditions of Theorem 3.2.1 or Theorem 3.2.2,*

$$\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}\right|=o_P((n\alpha)^{1/2}),\quad \sqrt{n\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}-1\right)\xrightarrow{\mathbb{P}}0,\quad \frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}-1\right)^2-\sigma_\alpha^2\xrightarrow{\mathbb{P}}0.$$

Proof. Note that we can write, for $i=1,\ldots,n$,

$$\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}=\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}+\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha},$$

and therefore

$$\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}=\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}.\quad (3.32)$$

Hence,

$$\left(1-\frac{1}{n\alpha}\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|\right)\left|\frac{\alpha\widehat{V}_i}{\widehat{\rho}_\alpha}\right|\le\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|+\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|+\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|$$
$$+\frac{1}{n\alpha}\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|.$$

Lemma 3.6.6 implies that

$$\left(1-o_P((n\alpha)^{-1/2})\right)\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}\right|=o_P((n\alpha)^{1/2}),$$

and then the *first* claim follows.

Note that (3.32) also yields that

$$\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_i}-1=\left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}-1\right)+\left(\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-1\right)-\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-1\right)+\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}T_i \tag{3.33}$$

where

$$\begin{aligned}T_i=&\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}+\frac{1}{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^2-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\\=:&T_{i,1}+T_{i,2}-T_{i,3}-T_{i,4}.\end{aligned}$$

With Lemmas 3.6.5 and 3.6.8, applying Cauchy-Schwarz inequality yields that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n|T_{i,1}|=&\frac{1}{n^2}\sum_{i=1}^n\left|\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|\\\le&\frac{1}{n\alpha}\sqrt{\frac{\alpha}{n}\sum_{i=1}^n\left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right)^2\cdot\frac{\alpha}{n}\sum_{i=1}^n\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^2}=O_P((n\alpha)^{-1}).\end{aligned}$$

Analogously, we can also show that for all $j=2,3,4$, $\frac{1}{n}\sum_{i=1}^n|T_{i,j}|=O_P\left((n\alpha)^{-1}\right)$.

Therefore

$$\frac{1}{n}\sum_{i=1}^{n}|T_i| \le \sum_{j=1}^{4}\frac{1}{n}\sum_{i=1}^{n}|T_{i,j}| = O_P((n\alpha)^{-1}).$$

Recalling $\max_{1\le i\le n}\left|\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right| = o_P((n\alpha)^{-1/2})$, we have that

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}T_i\right| \le & \frac{1}{n}\sum_{i=1}^{n}|T_i| \max_{1\le i\le n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}\right| \\
= & O_P((n\alpha)^{-1})\cdot O_P(1) = o_P((n\alpha)^{-1/2}),
\end{aligned}$$

and then the *second* claim follows by applying Lemma 3.6.5 with (3.33).

Similarly, we have

$$\begin{aligned}
\frac{\alpha}{n}\sum_{i=1}^{n}\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}T_i^2 \le & \frac{\alpha}{n}\sum_{i=1}^{n}T_i^2 \max_{1\le i\le n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}\right| \\
\le & n\alpha\left(\frac{1}{n}\sum_{i=1}^{n}|T_i|\right)^2 \max_{1\le i\le n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}\right| \\
= & n\alpha\cdot O_P\left((n\alpha)^{-2}\right)\cdot O_P(1) = O_P\left((n\alpha)^{-1}\right) = o_P(1).
\end{aligned}$$

The *third* claim then follows from Lemma 3.6.8, using (3.33) again. □

Proof of Theorem 3.2.3. Set $m_i = m_i(\rho_\alpha) = \frac{\alpha\widehat{V}_{\rho,i}}{\rho_\alpha} - \alpha$, $i = 1,\ldots,n$, $m_n^* = \max_{1\le i\le n}|m_i|$, $\bar{m}_n = n^{-1}\sum_{i=1}^{n} m_i$, $S_n = n^{-1}\sum_{i=1}^{n} m_i^2$. Now Lemma 3.6.7 and Lemma 3.6.9 in conjunction with Theorem 3.2.1 and Theorem 3.2.2 imply that, as $n\to\infty$,

$$m_n^* = o_P\left((n\alpha)^{1/2}\right), \quad \sqrt{\frac{n}{\alpha}}\frac{\bar{m}_n}{\sigma_\alpha} \xrightarrow{d} N(0,1), \quad \text{and} \quad \frac{S_n}{\alpha\sigma_\alpha^2} \xrightarrow{P} 1. \tag{3.34}$$

When taking $\theta = \rho_\alpha$, equation (3.3) can be rewritten as

$$p_i = \frac{1}{n}\frac{1}{1+\widetilde{\lambda}m_i}$$

with $\widetilde{\lambda} = \lambda\rho_\alpha/\alpha$ and equation (3.4) can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{1+\widetilde{\lambda}m_i} = 0.$$

Now, following the proof of Theorem 2 in [58], statement (3.34) implies that $\widetilde{\lambda} = O_P((n\alpha)^{-1/2})$ and, furthermore,

$$\widetilde{\lambda} = S_n^{-1}\bar{m}_n + o_P((n\alpha)^{-1/2}).$$

Hence, by a Taylor expansion and again (3.34), we have, as $n \to \infty$,

$$\begin{aligned}
-2\log\widehat{\mathcal{R}}(\rho_\alpha) &= 2\sum_{i=1}^{n}\widetilde{\lambda}m_i - \sum_{i=1}^{n}\widetilde{\lambda}^2 m_i^2 + o_P(1) \\
&= nS_n^{-1}\bar{m}_n^2 + o_P(1) \\
&\xrightarrow{d} \chi^2(1).
\end{aligned}$$

$\square$

3.6.2 AR-GARCH/IGARCH Model

All the conditions of Theorem 3.3.1 are assumed throughout this subsection. Some intermediate lemmas are needed and in their proofs we denote $\widetilde{\boldsymbol{Z}}_t$ as the non-smoothed counterpart of $\boldsymbol{Z}_t$ given by

$$\begin{aligned}
\widetilde{\boldsymbol{Z}}_t(\theta_5,\boldsymbol{\nu}) = \Bigg(& w_{x,t}\frac{\partial l_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}, \mathbb{1}[\eta_{x,t}(\boldsymbol{\psi}_x)\le\theta_1] - 1 + \alpha, -\eta_{x,t}(\boldsymbol{\psi}_x)\mathbb{1}[\eta_{x,t}(\boldsymbol{\psi}_x)\le\theta_1] - \theta_3, \\
& w_{y,t}\frac{\partial l_{y,t}(\boldsymbol{\psi}_y)}{\partial\boldsymbol{\psi}_y}, \mathbb{1}[\eta_{y,t}(\boldsymbol{\psi}_y)\le\theta_2] - 1 + \alpha, -\eta_{y,t}(\boldsymbol{\psi}_y)\mathbb{1}[\eta_{y,t}(\boldsymbol{\psi}_y)\le\theta_2] - \theta_4, \\
& 2\alpha - 1 + \mathbb{1}[\eta_{x,t}(\boldsymbol{\psi}_y)\le\theta_1, \eta_{y,t}(\boldsymbol{\psi}_y)\le\theta_2] - \alpha\theta_5\theta_4/\theta_3\Bigg).
\end{aligned}$$

where

$$w_{x,t} = \left(\max\left\{1, C_x^{-1}\sum_{k=1}^{\infty}\frac{|X_{t-k}|I(|X_{t-k} > C_x|)}{k^9}\right\}\right)^{-4},$$
$$w_{y,t} = \left(\max\left\{1, C_y^{-1}\sum_{k=1}^{\infty}\frac{|Y_{t-k}|I(|Y_{t-k} > C_y|)}{k^9}\right\}\right)^{-4}.$$

Here, the stationary sequences $\{w_{x,t}\}$ and $\{w_{y,t}\}$ are the population analogue of $\{\delta_{x,t}\}$, $\{\delta_{y,t}\}$. They are introduced mainly for a proper definition of Σ in Lemma 3.6.15 below. In practice we do not have the initial values X_t and Y_t when $t \leq 0$ and hence we always use weights $\{\delta_{x,t}\}$ and $\{\delta_{y,t}\}$ when constructing $\boldsymbol{Z}_t$; see (3.8).

Define $\xi_{x\rho t} = 1 + \sum_{i=0}^{\infty}\rho^i|\varepsilon_{xt-i}|$ and $\xi_{y\rho t}$ similarly, $t \in \mathbb{Z}$, for any $\rho \in (0,1)$. For all $a > 0$, we define the local parameter spaces $\Theta_{an}^x = \{\boldsymbol{\psi}_x : \|\boldsymbol{\psi}_x - \boldsymbol{\psi}_x^0\| \leq n^{-1/a}\}$, $\Theta_{an}^y = \{\boldsymbol{\psi}_y : \|\boldsymbol{\psi}_y - \boldsymbol{\psi}_y^0\| \leq n^{-1/a}\}$ and $\Theta_{an} = \{\boldsymbol{\nu} : \|\boldsymbol{\nu} - \boldsymbol{\nu}^0\| \leq n^{-1/a}\}$.

Lemma 3.6.10. *For all $\iota \in (0,1)$ and $2 \leq a < 2 + \min\{2(1-\iota)/\iota, \delta_0/2\}$*

$$\max_{1\leq i\leq n}\eta_{xt}^k\xi_{x\rho t-1}^{\iota} = o(n^{1/a}), \quad k = 0,1,2. \tag{3.35}$$

Similar statements hold with y substituting x everywhere.

Proof. Note that $E(\eta_t^k\xi_{x\rho t-1}^{\iota})^a = E(\eta_t^{ak})E(\xi_{x\rho t-1}^{a\iota}) < \infty$, the rest follows by a standard Borel-Cantelli argument; see, e.g., the first part of Lemma 3 in [43] and note that the independence requirement therein is not necessary. □

Lemma 3.6.11. *There exists constants $C > 0, \rho, \iota \in (0,1)$, independent of t and n, such* *that uniformly in $\boldsymbol{\psi}_x \in \Theta_{an}^x$ and $t = 1, \ldots, n$*

$$\left|\frac{h_{x,t}(\boldsymbol{\psi}_x)}{h_{x,t}} - 1\right| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1}, \tag{3.36}$$

$$\left|h_{x,t}^{-1/2}\varepsilon_{x,t}(\boldsymbol{\phi}_x)\right| \le C\xi^{\iota}_{x\rho t-1}, \quad \left|h_{x,t}^{-1/2}\frac{\partial\varepsilon_{x,t}(\boldsymbol{\phi}_x)}{\partial\boldsymbol{\psi}_x}\right| \le C\xi^{\iota}_{x\rho t-1}, \tag{3.37}$$

$$\left\|\frac{1}{h_{x,t}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \le C\xi^{\iota}_{x\rho t-1}, \quad \frac{1}{h_{x,t}}\left\|\frac{\partial^2 h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x\partial\boldsymbol{\psi}_x^T}\right\| \le C\xi^{\iota}_{x\rho t-1}, \tag{3.38}$$

$$\left|h_{x,t}^{-1/2}(\varepsilon_{x,t}(\boldsymbol{\phi}_x) - \varepsilon_{x,t})\right| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1}, \quad \frac{1}{h_{x,t}}\left\|\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x} - \frac{\partial h_{x,t}(\boldsymbol{\psi}_x^0)}{\partial\boldsymbol{\psi}_x^0}\right\| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1} \tag{3.39}$$

with probability 1 when n is large and $a \in (2, 2 + \min\{2(1-\iota)/\iota, \delta_0/2\})$. Similar statements hold with y substituting x everywhere.

Proof. The inequality below hold with probability 1 uniformly in $t = 1, \ldots, n$ when n is sufficiently large and ρ sufficiently close to 1. For presentation convenience we do not repeat this statement and the constant C may be different in different inequalities.

We first prove (3.36), from which (3.37) and (3.38) follow immediately in conjunction with Lemma 3.6.10 here and Lemma A.2, A.3 and A.5 in [70].

Applying Lemma A.2 and A.5 in [70] and Taylor expansion yields that

$$\sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x}\left|\log\frac{h_{x,t}(\boldsymbol{\psi}_x)}{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)}\right| \le n^{-1/a}\sup_{\boldsymbol{\psi}_x \in \Theta_n^x}\left\|\frac{1}{h_{x,t}(\boldsymbol{\psi}_x)}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\phi}_x}\right\| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1} = o(1).$$

Noting that $\lim_{x\downarrow 1}\frac{\log(x)}{x-1} = 1$, it follows

$$\sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x}\left|\frac{h_{x,t}(\boldsymbol{\psi}_x)}{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)} - 1\right| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1}. \tag{3.40}$$

Similarly, by Lemma A.3 and A.5 in [70] we have

$$\left|\left(\frac{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)}{h_{x,t}}\right)^{1/2} - 1\right| \le \frac{n^{-1/a}}{2h_{x,t}^{1/2}}\sup_{\boldsymbol{\psi}_x \in \Theta_n^x}\left\|\frac{1}{h_{x,t}^{1/2}(\boldsymbol{\psi}_x)}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\phi}_x}\right\| \le Cn^{-1/a}\xi^{\iota}_{x\rho t-1} = o(1), \tag{3.41}$$

and therefore

$$\sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x} \left| \frac{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)}{h_{x,t}} - 1 \right| \leq C n^{-1/a} \xi_{x\rho t-1}^{\iota} = o(1). \tag{3.42}$$

Now, combining (3.40) and (3.42) yields

$$\begin{aligned}\sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x} \left| \frac{h_{x,t}(\boldsymbol{\psi}_x)}{h_{x,t}} - 1 \right| \leq & \sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x} \left| \frac{h_{x,t}(\boldsymbol{\psi}_x)}{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)} - 1 \right| \sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x} \left| \frac{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)}{h_{x,t}} \right| \\ & + \sup_{\boldsymbol{\psi}_x \in \Theta_{an}^x} \left| \frac{h_{x,t}(\boldsymbol{\phi}_x, \boldsymbol{\phi}_{hx}^0)}{h_{x,t}} - 1 \right| \leq C n^{-1/a} \xi_{x\rho t-1}^{\iota}.\end{aligned}$$

This completes the proof of (3.36). Furthermore, by Taylor expansion we have

$$\sup_{\boldsymbol{\nu} \in \Theta_{an}} \left| h_{x,t}^{-1/2} (\varepsilon_{x,t}(\boldsymbol{\phi}_x) - \varepsilon_{x,t}) \right| \leq n^{-1/a} \sup_{\boldsymbol{\nu} \in \Theta_{an}} \left| h_{x,t}^{-1/2} \frac{\partial \varepsilon_{x,t}(\boldsymbol{\phi}_x)}{\partial \boldsymbol{\psi}_x} \right| \leq C n^{-1/a} \xi_{x\rho t-1}^{\iota},$$

where the last step follows from (3.37). The proof for the second part of (3.39) is analogous using (3.38). □

From now on we always take ρ and ι as those in Lemma 3.6.11.

Corollary 3.6.1. *Let $a \in (2, 2 + \min\{2(1-\iota)/\iota, \delta_0/2\})$. With some constant $C > 0$, independent of t and n, we have uniformly in $\boldsymbol{\psi}_x \in \Theta_{an}^x$ and $t = 1, \ldots, n$*

$$\begin{gathered}|\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}| \leq C n^{-1/a} (1 + |\eta_{x,t}|) \xi_{x\rho t-1}^{\iota}, \quad \left\| \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} \right\| \leq C (1 + |\eta_{x,t}|) \xi_{x\rho t-1}^{\iota}, \\ \left\| \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} - \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x^0} \right\| \leq C n^{-1/a} (1 + |\eta_{x,t}|) \xi_{x\rho t-1}^{2\iota}\end{gathered}$$

with probability 1 when n is large. Similar statements hold with y substituting x everywhere.

Proof. From Lemma 3.6.11 we have

$$\begin{aligned}|\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}| \leq & \left| \frac{h_{x,t}^{1/2}}{h_{x,t}^{1/2}(\boldsymbol{\psi}_x)} \right| \left| h_{x,t}^{-1/2} (\varepsilon_{x,t}(\boldsymbol{\phi}_x) - \varepsilon_{x,t}) \right| + |\eta_{x,t}| \left| \frac{h_{x,t}^{1/2}}{h_{x,t}^{1/2}(\boldsymbol{\psi}_x)} - 1 \right| \\ \leq & C n^{-1/a} (1 + |\eta_{x,t}|) \xi_{x\rho t-1}^{\iota}.\end{aligned}$$

Together with Lemma 3.6.11, it follows that

$$\begin{aligned}\left\|\frac{\partial\eta_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \le & \left|\frac{h_{x,t}^{1/2}}{h_{x,t}^{1/2}(\boldsymbol{\psi}_x)}\right|\left\|\frac{1}{h_{x,t}^{1/2}}\frac{\partial\varepsilon_{x,t}(\boldsymbol{\phi}_x)}{\partial\boldsymbol{\psi}_x}\right\| + \frac{1}{2}\left|\eta_{x,t}(\boldsymbol{\psi}_x)-\eta_{x,t}\right|\left|\frac{h_{x,t}}{h_{x,t}(\boldsymbol{\psi}_x)}\right|\left\|\frac{1}{h_{x,t}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \\ & + \frac{1}{2}|\eta_{x,t}|\left|\frac{h_{x,t}}{h_{x,t}(\boldsymbol{\psi}_x)}\right|\left\|\frac{1}{h_{x,t}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \\ \le & C\xi_{x\rho t-1}^{\iota} + Cn^{-1/a}(1+|\eta_{x,t}|)\xi_{x\rho t-1}^{2\iota} + C|\eta_{x,t}|\xi_{x\rho t-1}^{\iota} \le 2C(1+|\eta_{x,t}|)\xi_{x\rho t-1}^{\iota}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\left\|\frac{\partial\eta_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x} - \frac{\partial\eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial\boldsymbol{\psi}_x}\right\| \\ \le & \left\|\frac{h_{x,t}^{1/2}}{h_{x,t}^{1/2}(\boldsymbol{\psi}_x)}-1\right\|\cdot\frac{1}{h_{x,t}^{1/2}}\left\|\frac{\partial\varepsilon_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| + \left|\eta_{x,t}(\boldsymbol{\psi}_x)-\eta_{x,t}\right|\cdot\frac{h_{x,t}}{h_{x,t}(\boldsymbol{\psi}_x)}\cdot\left\|\frac{1}{h_{x,t}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \\ & + |\eta_{x,t}|\frac{h_{x,t}}{h_{x,t}(\boldsymbol{\psi}_x)}\cdot\frac{1}{h_{x,t}}\left\|\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x} - \frac{\partial h_{x,t}(\boldsymbol{\psi}_x^0)}{\partial\boldsymbol{\psi}_x^0}\right\| + \left\|\frac{h_{x,t}}{h_{x,t}(\boldsymbol{\psi}_x)}-1\right\|\cdot\left\|\frac{1}{h_{x,t}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\| \\ \le & Cn^{-1/a}\xi_{x\rho t-1}^{2\iota} + Cn^{-1/a}(1+|\eta_{x,t}|)\xi_{x\rho t-1}^{2\iota} + Cn^{-1/a}|\eta_{x,t}|\xi_{x\rho t-1}^{\iota} + n^{-1/a}\xi_{x\rho t-1}^{2\iota} \\ \le & Cn^{-1/a}(1+|\eta_{x,t}|)\xi_{x\rho t-1}^{2\iota}.\end{aligned}$$

□

Lemma 3.6.12. *Let* $a \in (2, 2+\min\{2(1-\iota)/\iota, \delta_0/2\})$. *With probability 1, as* $n\to\infty$,

$$\max_{1\le t\le n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\| = o(n^{1/a}). \tag{3.43}$$

Proof. It suffices to show the statement coordinate-wisely, i.e.

$$\max_{1\le t\le n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\left\|Z_{it}(\theta_5^0,\boldsymbol{\nu})\right\| = o(n^{1/a}), \quad i=1,\ldots,7, \tag{3.44}$$

uniformly in $\boldsymbol{\nu}\in\Theta_{an}$. This is trivial for $i=2,4,7$ since $|Z_{it}(\theta_5,\boldsymbol{\nu})|$ is bounded. In the sequel we shall only prove the statement for $i=1,3$; the proofs for $i=4,6$ are completely analogous and therefore omitted.

From Lemmas 3.6.10, 3.6.11 and Corollary 3.6.1, we know there exists a constant $C > 0$, independent of t and n, such that with probability 1

$$\begin{aligned}\left\|Z_{1t}(\theta_5^0, \boldsymbol{\nu})\right\| = \left\|\delta_{x,t}\frac{\partial l_{x,t}(\theta_5^0, \boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| \leq & C|\eta_{x,t}(\boldsymbol{\psi}_x)| \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| + C\frac{1}{h_{x,t}(\boldsymbol{\psi}_x)}\left\|\frac{\partial h_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| \\ \leq & 2C(\eta_{x,t}^2 + |\eta_{x,t}| + 1)\xi^{\iota}_{x\rho t-1} = o(n^{1/a})\end{aligned}$$

and, furthermore,

$$\left|Z_{3t}(\boldsymbol{\nu}, \theta^1, \theta_3)\right| \leq |\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}| + |\eta_{x,t}| + |\theta_3| = o(n^{1/a})$$

uniformly in $\boldsymbol{\psi}_x \in \Theta^x_{an}$ and $t = 1, \ldots, n$ when n is large. □

Lemma 3.6.13. *Let* $a \in (2, 2 + \min\{2(1-\iota)/\iota, \delta_0/2, \delta_h\})$. *As* $n \to \infty$

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\psi}_x \in \Theta^x_{an}}\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right\|\xi^{\iota}_{x\rho t-1} \xrightarrow{P} 0, \tag{3.45}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\psi}_y \in \Theta^y_{an}}\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{y,t}(\boldsymbol{\psi}_y)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{y,t}}{h}\right)\right\|\xi^{\iota}_{y\rho t-1} \xrightarrow{P} 0. \tag{3.46}$$

Proof. We only prove the first statement; the proof of the second one is completely analogous. All the inequalities below hold uniformly in $t = 1, \ldots, n$ and large n; for presentation convenience, we do not repeat this statement and the constant C may be different in different inequalities.

Write

$$\begin{aligned}&\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right\|\xi^{\iota}_{x\rho t-1} \\ =&\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right\|\mathbb{1}\left[\left|\eta_{x,t} - \theta_1^0\right| < h\right]\xi^{\iota}_{x\rho t-1} \\ &+\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right)\right\|\mathbb{1}\left[h \leq \left|\eta_{x,t} - \theta_1^0\right| \leq h + n^{-1/a} + |\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}|\right]\xi^{\iota}_{x\rho t-1} \\ =&T_{1t} + T_{2t}.\end{aligned}$$

Noting that $\sup_s |k'(s)| < \infty$, by Taylor expansion it is easy to show

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} T_{1t} \leq & \frac{1}{n}\sum_{t=1}^{n} \frac{C}{n^{1/a}h^2}\left(1+\sup_{\boldsymbol{\psi}_x\in\Theta_{an}^x}\left\|\frac{\partial\eta_{x,t}(\boldsymbol{\psi}_x)}{\partial\boldsymbol{\psi}_x}\right\|\right)\mathbb{1}\left[|\eta_{x,t}-\theta_1^0|<h\right]\xi_{x\rho t-1}^{\iota} \\ \leq & \frac{1}{n}\sum_{t=1}^{n} \frac{C}{n^{1/a}h^2}\left(1+C(|\eta_{x,t}|+1)\xi_{x\rho t-1}^{\iota}\right)\mathbb{1}\left[|\eta_{x,t}-\theta_1^0|\leq h\right]\xi_{x\rho t-1}^{\iota} \\ \leq & \frac{1}{n}\sum_{t=1}^{n} \frac{2C}{n^{1/a}h^2}\xi_{x\rho t-1}^{2\iota}\mathbb{1}\left[|\eta_{x,t}-\theta_1^0|\leq h\right] = o_p(1)\end{aligned}$$

where the last step results from Markov inequality and the fact that

$$E(\xi_{x\rho t-1}^{\iota}\mathbb{1}\left[|\eta_{x,t}-\theta_1^0|\leq h\right]) = E(\xi_{x\rho t-1}^{2\iota})P(|\eta_{x,t}-\theta_1^0|\leq h) = O(h) = o(n^{1/a}h^2).$$

Recall $\max_{1\leq t\leq n}(|\eta_{x,t}|+1)\xi_{x\rho t-1}^{\iota} = o(n^{1/a})$ from Lemma 3.6.10 and $\sup_s k(s) < \infty$. For large n, we then have with probability 1

$$\begin{aligned}T_{2t} \leq & \frac{C}{h}\mathbb{1}\left[h\leq|\eta_{x,t}-\theta_1^0|\leq h+n^{-1/a}+Cn^{-1/a}(|\eta_{x,t}|+1)\xi_{x\rho t-1}^{\iota}\right]\xi_{x\rho t-1}^{\iota} \\ \leq & \frac{C}{h}\mathbb{1}\left[h\leq|\eta_{x,t}-\theta_1^0|\leq h+n^{-1/a}+C^2n^{-1/a}\xi_{x\rho t-1}^{\iota}\right]\xi_{x\rho t-1}^{\iota} =: T_{2t}^{+}.\end{aligned}$$

Applying an appropriate weak law of large number for triangular arrays, e.g., Theorem 1 in [79], we can show

$$\frac{1}{n}\sum_{t=1}^{n} T_{2t}^{+} = \frac{1}{n}\sum_{t=1}^{n} E(T_{2t}^{+}\mathbb{1}[T_{2t}\leq n]|\mathcal{F}_{t-1}) + o_p(1). \tag{3.47}$$

Furthermore, for large n with probability 1

$$
\begin{aligned}
0 \leq & \frac{1}{n}\sum_{t=1}^{n} E(T_{2t}^{+}\mathbb{1}[T_{2t} \leq n]|\mathcal{F}_{t-1}) \\
\leq & \frac{C}{nh}\sum_{t=1}^{n} \xi_{x\rho t-1}\Big\{\Big(F_{\eta_{x,t}}(\theta_1^0 + h + n^{-1/a} + n^{-1/a}\xi_{x\rho t-1}^{\iota}) - F_{\eta_{x,t}}(\theta_1^0 + h)\Big) \\
& + \Big(F_{\eta_{x,t}}(\theta_1^0 - h) - F_{\eta_{x,t}}(\theta_1^0 - h - n^{-1/a} - n^{-1/a}\xi_{x\rho t-1}^{\iota})\Big)\Big\} \\
\leq & \frac{C}{nh}\sum_{t=1}^{n} \xi_{x\rho t-1}^{\iota}(n^{-1/a} + n^{-1/a}\xi_{x\rho t-1}^{\iota}) = \frac{2C}{n^{1/a}h}\frac{1}{n}\sum_{t=1}^{n}\xi_{x\rho t-1}^{2\iota} = o_p(1)
\end{aligned}
$$

where the last equality results from Markov inequality and the fact that $E(\xi_{x\rho t-1}^{2\iota}) = O(1) = o(n^{1/a}h)$. Hence, we have $\frac{1}{n}\sum_{t=1}^{n} T_{2t} = o_p(1)$, and this completes the proof. □

Lemma 3.6.14. *Let $a \in (2, 2+\min\{2(1-\iota)/\iota, \delta_0/2, \delta_h\})$. As $n \to \infty$,*

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} = V + o_p(1) \tag{3.48}
$$

uniformly in $\boldsymbol{\nu} \in \Theta_{an}$ and the matrix V is given in appendix.

Proof. Using (A.19) in [70] and a weak law of large number for triangular arrays, e.g. Theorem 1 in [79], it can be easily shown that

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0)}{\partial \boldsymbol{\nu}} \xrightarrow{P} V.
$$

For the rest it suffices to show

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial Z_{it}(\theta_5, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial Z_{it}(\theta_5^0, \boldsymbol{\nu}^0)}{\partial \boldsymbol{\nu}} = o_p(1), \quad i = 1, \ldots, 7, \tag{3.49}
$$

uniformly in $\boldsymbol{\nu} \in \Theta_{an}$. This is readily known for $i = 1, 4$, recalling (A.19) in [70] again. In the sequel we shall only prove (3.49) for $i = 3$ and $i = 7$; the proof for other cases are analogous and therefore omitted. All the statements below hold uniformly in $\boldsymbol{\nu} \in \Theta_{an}$ and we do not repeat this argument for presentation convenience. The constant $C > 0$ below may be different in different inequalities.

Write

$$\frac{\partial Z_{3t}(\boldsymbol{\psi}_x, \theta_1)}{\partial \boldsymbol{\psi}_x} = -\frac{1}{h} k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) \eta_{x,t}(\boldsymbol{\psi}_x) \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} + \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} K\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right).$$

Note that, when $|\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)| \leq h$ or $|\theta_1^0 - \eta_{x,t}| \leq h$, we have

$$|\theta_1^0 - \eta_{x,t}(\boldsymbol{\psi}_x)| \leq |\theta_1 - \theta_1^0| + h + |\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}| = o(1),$$

i.e. $\eta_{x,t}(\boldsymbol{\psi}_x) = \theta_1^0 + o(1)$, implying that

$$\left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| \leq C\xi^{\iota}_{x\rho t-1}, \quad \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} - \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right\| \leq Cn^{-1/a}\xi^{\iota}_{x\rho t-1}$$

uniformly in $t = 1, \ldots, n$ with probability 1. It follows that, for large n

$$\begin{aligned}
&\left\|\frac{\partial Z_{3t}(\boldsymbol{\psi}_x, \theta_1)}{\partial \boldsymbol{\psi}_x} - \frac{\partial Z_{3t}(\boldsymbol{\psi}_x^0, \theta_1^0)}{\partial \boldsymbol{\psi}_x^0}\right\| \\
\leq & \left|\frac{1}{h} k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h} k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right| |\eta_{x,t}(\boldsymbol{\psi}_x)| \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| \\
&+ \frac{1}{h} k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right) \left\{|\eta_{x,t}(\boldsymbol{\psi}_x) - \eta_{x,t}| \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| + |\eta_{x,t}| \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} - \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right\|\right\} \\
&+ \left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} - \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right\| + \left|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right| \left|K\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - K\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right| \\
\leq & C|\theta_1^0| \left|\frac{1}{h} k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h} k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right| \xi^{\iota}_{x\rho t-1} + C|\theta_1^0| n^{-1/a} \frac{1}{h} k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right) \xi^{2\iota}_{x\rho t-1} \\
&+ C\xi^{\iota}_{x\rho t-1} \mathbb{1}[|\eta_{x,t} - \theta_1^0| \leq h] =: T_{1t} + T_{2t} + T_{3t}
\end{aligned}$$

with probability 1. We already know from Lemma 3.6.13 that $\frac{1}{n}\sum_{t=1}^n T_{1t} = o_p(1)$. Moreover, a standard argument using Markov inequality yields

$$\frac{1}{n}\sum_{t=1}^{n} T_{2t} = o_p(1), \quad \frac{1}{n}\sum_{t=1}^{n} T_{3t} = o_p(1)$$

with the facts that

$$E(T_{2t}) = C|\theta_1^0|n^{-1/a}(\theta_1^0 + (1))E(\xi_{x\rho t-1}^{2\iota}) = o(1)$$
$$E(T_{3t}) = CE(\xi_{x\rho t-1}^{\iota})P(|\eta_{x,t} - \theta_1^0| \le h) = o(h) = o(1).$$

It then follows from above equations that

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial Z_{3t}(\boldsymbol{\psi}_x,\theta_1)}{\partial \boldsymbol{\psi}_x} - \frac{\partial Z_{3t}(\boldsymbol{\psi}_x^0,\theta_1^0)}{\partial \boldsymbol{\psi}_x}\right\| = o_p(1).$$

The rest of (3.49) for $i = 3$ can be shown similarly.

Next, we shall prove the case for $i = 7$. For large n

$$\begin{aligned}
&\left\|\frac{\partial Z_{7t}(\theta_5^0,\boldsymbol{\nu})}{\partial \boldsymbol{\psi}_x} - \frac{\partial Z_{7t}(\theta_5^0,\boldsymbol{\nu}^0)}{\partial \boldsymbol{\psi}_x^0}\right\| \\
\le&\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right\|\left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x}\right\| \\
&+\frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\left\|\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x)}{\partial \boldsymbol{\psi}_x} - \frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right\| \\
&+\frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\frac{\partial \eta_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\left|K\left(\frac{\theta_2 - \eta_{y,t}(\boldsymbol{\psi}_y)}{h}\right) - K\left(\frac{\theta_2^0 - \eta_{y,t}}{h}\right)\right| \\
\le&C\left\|\frac{1}{h}k\left(\frac{\theta_1 - \eta_{x,t}(\boldsymbol{\psi}_x)}{h}\right) - \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)\right\|\xi_{x\rho t-1}^{\iota} + \frac{1}{h}k\left(\frac{\theta_1^0 - \eta_{x,t}}{h}\right)n^{-1/a}\xi_{x\rho t-1}^{\iota} \\
&+\frac{C}{h}\xi_{x\rho t-1}^{\iota}\mathbb{1}[|\theta_1^0 - \eta_{x,t}| \le h, |\theta_2^0 - \eta_{y,t}| \le h] =: T'_{1t} + T'_{2t} + T'_{3t}
\end{aligned}$$

with probability 1. As above, by Lemma 3.6.13 and by using Markov inequality with the facts that $E(T'_{2t}) = O(n^{-1/a}) = o(1)$ and $E(T'_{3t}) = O(h) = o(1)$, we have $\frac{1}{n}\sum_{t=1}^{n} T'_{1t} =$

$o_p(1)$, $\frac{1}{n}\sum_{t=1}^{n} T'_{2t} = o_p(1)$ and $\frac{1}{n}\sum_{t=1}^{n} T'_{3t} = o_p(1)$. Hence,

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial Z_{7t}(\theta_5^0,\boldsymbol{\nu})}{\partial \boldsymbol{\psi}_x} - \frac{\partial Z_{7t}(\theta_5^0,\boldsymbol{\nu}^0)}{\partial \boldsymbol{\psi}_x}\right\| = o_p(1).$$

The rest of (3.49) for $i = 7$ can be shown similarly. $\square$

Lemma 3.6.15. *Let $a \in (2, 2 + \min\{2(1-\iota)/\iota, \delta_0/2, \delta_h\})$. Uniformly in $\boldsymbol{\nu} \in \Theta_{an}$*

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})^T = \Sigma + o_p(1)$$

where $\Sigma := E(\widetilde{\boldsymbol{Z}}_t(\theta_5^0,\boldsymbol{\nu}^0)\widetilde{\boldsymbol{Z}}_t(\theta_5^0,\boldsymbol{\nu}^0)^T)$ is positive definite.

Proof. Using the law of large number for martingales (see, e.g., Theorem 2.13 in [42]), and Lemma A.4 in [70] (as in the last paragraph of the proof of Theorem 3.1 (ii) therein), we can show that

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)^T = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{Z}}_t(\theta_5^0,\boldsymbol{\nu}^0)\widetilde{\boldsymbol{Z}}_t(\theta_5^0,\boldsymbol{\nu}^0)^T + o_p(1) = \Sigma + o_p(1). \quad (3.50)$$

Now, for large n, by Lemmas 3.6.10, 3.6.11 and Corollary 3.6.1, we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})^T - \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)^T\right\| \\
\leq&\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})^T - \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)^T\right\| \\
\leq&2\max_{1\leq t\leq n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\|\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}) - \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)\| \\
\leq&2n^{-1/a}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\|\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\left\|\frac{\partial\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}\right\| \\
\leq&Cn^{-1/a}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\|\left\{\frac{1}{n}\sum_{t=1}^{n}(1+|\eta_{x,t}|)^2\xi_{x\rho t-1}^{2\iota} + \frac{1}{n}\sum_{t=1}^{n}(1+|\eta_{y,t}|)^2\xi_{y\rho t-1}^{2\iota}\right\} \\
=&o_p(1)\cdot O_p(1) = o_p(1)
\end{aligned}$$

with probability 1, where $C > 0$ is some constant independent of t and n. The lemma then follows; the positive definiteness Σ can be easily verified using Assumption A4. $\square$

Lemma 3.6.16. *As* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0) \xrightarrow{d} N(0, \Sigma). \tag{3.51}$$

Proof. Using the fact that k is symmetric around the origin, it is easy to show that

$$|E(\boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0)) - E(\widetilde{\boldsymbol{Z}}_t(\theta_5^0, \boldsymbol{\nu}^0))| = |E(\boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0))| = O(h^2) = o(n^{-1/2}). \tag{3.52}$$

By Markov inequality we have

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0) - \frac{1}{n} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_t(\theta_5^0, \boldsymbol{\nu}^0) = o_p(n^{-1/2}). \tag{3.53}$$

Hence, recalling (3.50) and applying the martingale central limit theorem (see, e.g., Corollary 3.1 in [42]) yields that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}^0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\boldsymbol{Z}}_t(\theta_5^0, \boldsymbol{\nu}^0) + o_p(1) \xrightarrow{d} N(0, \Sigma), \quad n \to \infty. \tag{3.54}$$

$\square$

Lemma 3.6.17. *Let* $a \in (2, 2 + \min\{2(1-\iota)/\iota, \delta_0/2, \delta_h, 1\})$. *Then*

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \boldsymbol{Z}_t(\theta_5^0, \boldsymbol{\nu}) \right\|^3 = o_p(n^{3/a-1}) \tag{3.55}$$

uniformly in $\boldsymbol{\nu} \in \Theta_{an}$.

Proof. Our proof is very similar to that of Lemma 11.3 in [1]. By Lemma 3.6.12 and

Markov inequality, we know uniformly in $\boldsymbol{\nu} \in \Theta_{an}$ for large n

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\nu}\in\Theta_{an}}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|^a \leq \frac{C}{n}\sum_{t=1}^{n}(1+|\eta_{x,t}|)^a\xi_{x\rho t-1}^{a\iota} + \frac{C}{n}\sum_{t=1}^{n}(1+|\eta_{y,t}|)^a\xi_{y\rho t-1}^{a\iota} = O_p(1) \tag{3.56}$$

where $C > 0$ is some constant independent of t and n. It follows that

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|^3 \leq \frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|^a \max_{1\leq t\leq n}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|^{3-a} = O_p(1)\cdot o_p(n^{(3-a)/a})$$

uniformly in $\boldsymbol{\nu} \in \Theta_{an}$, as $n \to \infty$. □

Lemma 3.6.18. *Let* $a \in (2, 2+\min\{2(1-\iota)/\iota, \delta_0/2, \delta_h, 1\})$. *When* $\boldsymbol{\nu}_n \in \Theta_{an}$*,*

$$-2\log L(\theta_5^0,\boldsymbol{\nu}_n) = \boldsymbol{\theta}^T V^T\Sigma^{-1}V\boldsymbol{\theta}+2\boldsymbol{\theta}^T V^T\Sigma^{-1}\mathbb{Z}_n+\mathbb{Z}_n\Sigma^{-1}\mathbb{Z}_n+o_p(1)+o_p(\|\boldsymbol{\theta}\|)+o_p(\|\boldsymbol{\theta}\|^2) \tag{3.57}$$

with $\mathbb{Z}_n := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(\theta_5^0,\nu^0)$, $\boldsymbol{\theta} := \sqrt{n}(\boldsymbol{\nu}_n - \boldsymbol{\nu}^0)$*, and* V *and* Σ *from Lemmas 3.6.14 and 3.6.15.*

Proof. Many arguments below are similar to those in Owen (1990) and those in the proof of Lemma 1 in [2]. We refer many of details below to these two seminal papers.

With Lemmas 3.6.14 and 3.6.15, similar to the proof of Owen (1990) we can show that

$$\boldsymbol{\lambda}(\boldsymbol{\nu}) = O_p(n^{-1/a}) \tag{3.58}$$

uniformly in $\boldsymbol{\nu} \in \Theta_{an}$; see also the proof of (A.1) in [2]. It follows that, further in conjunction with Lemma 3.6.17, we may write

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{\nu}) =& \Sigma^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0) + V(\boldsymbol{\nu}-\boldsymbol{\nu}^0)\right) + o_P(n^{-1/2}) + o_p(\left\|\boldsymbol{\nu}-\boldsymbol{\nu}^0\right\|)\\ =& n^{-1/2}\Sigma^{-1}(\mathbb{Z}_n + V\boldsymbol{\theta}) + o_p(n^{-1/2}) + o_p(n^{-1/2}\|\boldsymbol{\theta}\|)\end{aligned} \tag{3.59}$$

as in the proof of (2.17) in [43].

Now we may expand

$$
\begin{aligned}
&-2\log L(\theta_5^0,\boldsymbol{\nu})\\
=&2n\boldsymbol{\lambda}(\boldsymbol{\nu})^T\frac{1}{n}\sum_{t=1}^n \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}) - n\cdot\frac{1}{n}\sum_{t=1}^n\left(\boldsymbol{\lambda}(\boldsymbol{\nu})^T\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right)^2+\frac{2n}{3}\frac{1}{n}\sum_{t=1}^n\frac{(\boldsymbol{\lambda}(\boldsymbol{\nu})^T\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}))^3}{(1+\delta_t)}\\
=:&S_1-S_2+S_3
\end{aligned}
$$

where, by Lemma 3.6.10 and (2.15),

$$
\max_{1\le t\le n}|\delta_t|\le\|\boldsymbol{\lambda}(\boldsymbol{\nu})\|\max_{1\le t\le n}\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|=O_p(n^{-1/a})\cdot o_p(n^{-1/a})=o_p(1)
$$

and therefore S_3 has a norm bounded by

$$
n\cdot\|\boldsymbol{\lambda}(\boldsymbol{\nu})\|^3\frac{1}{n}\sum_{t=1}^n\left\|\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\right\|^3\max_{1\le t\le n}\frac{1}{|1+\delta_t|}=n\cdot O_p(n^{-3/a})\cdot o_p(n^{3/a-1})O_p(1)=o_p(1).
$$

Write

$$
S_1=2n^{1/2}\boldsymbol{\lambda}(\boldsymbol{\nu})^T\mathbb{Z}_n+2n\boldsymbol{\lambda}(\boldsymbol{\nu})^T\left\{\frac{1}{n}\sum_{t=1}^n\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})-\frac{1}{n}\sum_{t=1}^n\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}^0)\right\}=:S_{11}+S_{12}.
$$

Substituting (2.16) in above equation yields that

$$
S_{11}=2\mathbb{Z}_n^T\Sigma^{-1}\mathbb{Z}_n+2\boldsymbol{\theta}^TV^T\Sigma^{-1}\mathbb{Z}_n+o_p(1)+o_p(\|\boldsymbol{\theta}\|)
$$

and, together with Lemma 3.6.14, for some $\boldsymbol{\nu}_n^*\in\Theta_{an}$,

$$
\begin{aligned}
S_{12}=&2n\boldsymbol{\lambda}(\boldsymbol{\nu})^T\frac{1}{n}\sum_{t=1}^n\frac{\partial\boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu}_n^*)}{\partial\boldsymbol{\nu}}(\boldsymbol{\nu}-\boldsymbol{\nu}^0)\\
=&2n\boldsymbol{\lambda}(\boldsymbol{\nu})(V+o_p(1))(\boldsymbol{\nu}-\boldsymbol{\nu}^0)\\
=&2\boldsymbol{\theta}^TV^T\Sigma^{-1}\mathbb{Z}_n+2\boldsymbol{\theta}^TV^T\Sigma^{-1}V\boldsymbol{\theta}+o_p(\|\boldsymbol{\theta}\|)+o_p(\|\boldsymbol{\theta}\|^2).
\end{aligned}
$$

Similarly, substituting (2.16) and using Lemma 3.6.15 yields that

$$\begin{aligned}S_2 =&\boldsymbol{\lambda}(\boldsymbol{\nu})^T \left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t(\theta_5^0,\boldsymbol{\nu})\boldsymbol{Z}_t^T(\theta_5^0,\boldsymbol{\nu})\right]\boldsymbol{\lambda}(\boldsymbol{\nu})\\ =&\boldsymbol{\lambda}(\boldsymbol{\nu})^T(\Sigma+o_p(1))\boldsymbol{\lambda}(\boldsymbol{\nu})\\ =&\mathbb{Z}_n^T\Sigma^{-1}\mathbb{Z}_n+2\boldsymbol{\theta}^T V^T\Sigma^{-1}\mathbb{Z}_n+\boldsymbol{\theta}^T V^T\Sigma^{-1}V\boldsymbol{\theta}+o_p(1)+o_p(\|\boldsymbol{\theta}\|)+o_p(\|\boldsymbol{\theta}\|^2).\end{aligned}$$

The lemma then follows. □

Lemma 3.6.19. *Let $a \in (2, 2+\min\{2(1-\iota)/\iota, \delta_0/2, \delta_h, 1\})$. As $n \to \infty$, with probability tending to 1, $-2\log L(\theta_5^0, \boldsymbol{\nu})$ obtains its minimal value at some point $\widetilde{\boldsymbol{\nu}}$ in the interior of the ball $\|\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}^0\| \leq n^{-1/a}$ and, furthermore,*

$$\sqrt{n}(\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}^0) = -\left(V^T\Sigma^{-1}V\right)^{-1}V^T\Sigma^{-1}\mathbb{Z}_n+o_p(1), \tag{3.60}$$

where V, Σ and $\mathbb{Z}_n$ are from Lemmas 3.6.14, 3.6.15 and 3.6.18.

Proof. The proof of the first part is completely analogous to that of Lemma 1 in [2] using Lemmas 3.6.14, 3.6.15 and 3.6.16 and therefore omitted. The proof of the second part below is a (slight) modification of that of Theorem 2 in [44].

By definition we have

$$-2\log L(\theta_5^0, \widetilde{\boldsymbol{\nu}}) \leq -2\log L(\theta_5^0, \boldsymbol{\nu}^0).$$

Denote $\sqrt{n}(\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}^0) = \widetilde{\boldsymbol{\theta}}$ and recall from Lemma 3.6.16 that $\mathbb{Z}_n = O_p(1)$. Applying (3.57) twice in the last expression, consolidating terms, and using the fact that Σ is positive definite (therefore $V^T\Sigma^{-1}V$ is positive definite), we get

$$\begin{aligned}0 \leq& \left(\widetilde{\boldsymbol{\theta}}+\left(V^T\Sigma^{-1}V\right)^{-1}V^T\Sigma^{-1}\mathbb{Z}_n\right)V^T\Sigma^{-1}V\left(\widetilde{\boldsymbol{\theta}}+\left(V^T\Sigma^{-1}V\right)^{-1}V^T\Sigma^{-1}\mathbb{Z}_n\right)\\ \leq& o_p(1)+o_p(\|\widetilde{\boldsymbol{\theta}}\|)+o_p(\|\widetilde{\boldsymbol{\theta}}\|^2) = o_p(\|\widetilde{\boldsymbol{\theta}}+\left(V^T\Sigma^{-1}V\right)^{-1}V^T\Sigma^{-1}\mathbb{Z}_n\|^2).\end{aligned}$$

It follows that $\widetilde{\boldsymbol{\theta}} + \left(V^T\Sigma^{-1}V\right)^{-1} V^T\Sigma^{-1}\mathbb{Z}_n = o_p(1)$, that is

$$\sqrt{n}(\widetilde{\boldsymbol{\nu}} - \boldsymbol{\nu}^0) = -\left(V^T\Sigma^{-1}V\right)^{-1} V^T\Sigma^{-1}\mathbb{Z}_n + o_p(1). \tag{3.61}$$

This completes the proof. $\square$

Proof of Theorem 3.3.1. Recall V, Σ and $\mathbb{Z}_n$ from Lemmas 3.6.14, 3.6.15 and 3.6.18. From Lemma 3.6.19 we have that the maximum empirical likelihood estimator $\widetilde{\boldsymbol{\nu}} \in \Theta_{an}$ and

$$\sqrt{n}(\widetilde{\boldsymbol{\nu}} - \boldsymbol{\nu}^0) = -\left(V^T\Sigma^{-1}V\right)^{-1} V^T\Sigma^{-1}\mathbb{Z}_n + o_p(1).$$

Substituting this into (3.57) yields that

$$-2\log L^P(\theta_5^0) = -2\log L(\theta_5^0, \widetilde{\boldsymbol{\nu}}) = \mathbb{Z}_n\Sigma^{-1/2}D\Sigma^{-1/2}\mathbb{Z}_n$$

where

$$D := \left\{I - \Sigma^{-1/2}V(V^T\Sigma^{-1}V)^{-1}V^T\Sigma^{-1/2}\right\}. \tag{3.62}$$

Note that $\Sigma^{-1/2}\mathbb{Z}_n \xrightarrow{d} N(0, I)$ by Lemma 3.6.16, and D is symmetric, idempotent and

$$\begin{aligned}\text{tr}(D) =& \text{dim}(\boldsymbol{\nu}^0) + 1 - tr(V^T\Sigma^{-1/2} \cdot \Sigma^{-1/2}V(V^T\Sigma^{-1}V)^{-1})\\ =& \text{dim}(\boldsymbol{\nu}^0) + 1 - \text{dim}(\boldsymbol{\nu}^0) = 1.\end{aligned}$$

The theorem then follows. $\square$

CHAPTER 4

A NEW TAIL DEPENDENCE MEASURE

Modeling and forecasting extreme co-movements in financial market is important for conducting stress test in risk management. Asymptotic independence and asymptotic dependence behave drastically different in modeling such co-movements. For example, the impact of extreme events is usually overestimated whenever asymptotic dependence is wrongly assumed. On the other hand, the impact is seriously underestimated whenever the data is misspecified as asymptotic independent. Therefore, distinguishing between asymptotic independence/dependence scenarios is very informative for any decision-making and especially in risk management. We investigate the properties of the limiting conditional Kendall's tau which can be used to detect the presence of asymptotic independence/dependence. We also propose nonparametric estimation for this new measure and derive its asymptotic limit. A simulation study shows the good performances of the new measure and its combination with the coefficient of tail dependence proposed by [53, 54]. Finally, applications to financial and insurance data are provided. The content of this chapter is based on A. Asimit, R. Gerrard, Y. Hou and L. Peng (2016). Tail Dependence Measure for Modeling Financial Extreme Co-movements. *Journal of Econometrics 194, 330-348.*

4.1 Introduction

An important task in risk management is to understand the reliability of the proposed model in the presence of adverse scenarios, known as stress testing. For example, the assessment of the capital adequacy in banking and insurance industries is based on quantifying the impact of extreme events on the solvability of financial and insurance conglomerates. Harmonized regulatory methodologies have been imposed in the banking industry (known as Basel III; see, [80]), and insurance industry within the European Union (known as Solvency

II; see, [81]) and in Switzerland (known as Swiss Solvency Test; see, [82]), that imposed the implementation of stress testing. It is generally accepted that Extreme Value Theory provides the appropriate technology to address the quantitative side of the problem (see for example, [83] and [84]). Since multiple sources of risks are competitive contributors to the calculations of the level of capital requirements, a holistic approach is to characterize such co-movements of extremes and then to effectively extrapolate data into tail region, which can naturally be done under the umbrella of Multivariate Extreme Theory as explained below.

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent and identically distributed random vectors with distribution function F and marginal distributions F_1 and F_2, i.e. $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. Bivariate Extreme Value Theory assumes that there are constants $a_n > 0, c_n > 0, b_n \in R, d_n \in R$ such that

$$\lim_{n\to\infty} \Pr\left(a_n\Big(\max_{1\le i\le n} X_i - b_n\Big) \le x, c_n\Big(\max_{1\le i\le n} Y_i - d_n\Big) \le y\right) = G(x, y), \tag{4.1}$$

for all continuous points (x, y) of G. In this case, G is called an *extreme value distribution* and F is said to belong to the *domain of attraction of* G. It follows from (4.1) that the following dependence convergence holds:

$$\begin{aligned} & \lim\nolimits_{t\to 0} t^{-1}\Big\{1 - F\big((1-F_1)^-(tx), (1-F_2)^-(ty)\big)\Big\} \\ = \quad & -\log G\big((-\log G_1)^-(x), (-\log G_2)^-(y)\big) \\ := \quad & l(x, y) \end{aligned} \tag{4.2}$$

for all $x, y \ge 0$, where $G_1(x) = G(x, \infty)$, $G_2(y) = G(\infty, y)$ and $(\cdot)^-$ denotes the left continuous inverse function. Here, $l(x, y)$ is called the tail dependence function (see [9]). It is easy to check that $l(ax, ay) = al(x, y)$ for all $a, x, y \ge 0$ and $x \vee y \le l(x, y) \le x + y$. This homogeneous property has been employed to extrapolate data into a tail region so that extreme events can be predicted (for details, see for example, [10]). However, when

$l(x, y) = x + y$, equation (4.2) implies that

$$\lim_{t \to 0} t^{-1} \Pr(1 - F_1(X_1) < tx, 1 - F_2(Y_1) < ty) = 0, \tag{4.3}$$

which makes extrapolation, i.e. statistical inference, impossible for concomitant extreme sets. In this case, F is said to have the asymptotic independence property, and a different convergence rate condition in (4.3) is needed for predicting joint extreme events. In other words, extreme value condition (4.1) is not enough for predicting extreme events in case of asymptotic independence. If the limit in (4.3) is not identical to zero, then F is said to have the *asymptotic dependence* property. It is known that a bivariate normal distribution with correlation coefficient between -1 and 1 is asymptotically independent, i.e. (4.3) holds (for details, see [11]).

Estimation of multivariate extreme becomes possible if the presence of asymptotic dependence/independece is known, and therefore, distinguishing between the two properties plays an important role in predicting extreme events. A mathematical formulation of this problem is made in [53, 54], where the coefficient of tail dependence, $0 < \eta \leq 1$, is introduced. It was assumed that

$$\Pr(1 - F_1(X_1) \leq t, 1 - F_2(Y_1) \leq t) = t^{1/\eta} s(t), \tag{4.4}$$

where $s(t)$ is a slowly varying function, i.e. $\lim_{t \to 0} s(tx)/s(t) = 1$ for all $x > 0$. Under condition (4.4), when $\eta = 1$ and $\lim_{t \to 0} s(t) = c \in (0, 1]$, the asymptotically dependent property holds, while either $\eta < 1$ or $\eta = 1$ and $\lim_{t \to 0} s(t) = 0$ implies asymptotic independence. Therefore, η and the limit behavior of function $s(t)$ can be used to distinguish between asymptotic dependence and asymptotic independence. Nonparametric inference for η can be found in [85] and [86]. Recently, [87] considered an asymptotically unbiased estimator for η in the case of $\eta < 1$, i.e. asymptotic independence.

Our proposal appeals to a robust measure of association that is appealing to a wide audience, and we find that most of the extreme scenarios are characterized by our method in order to elaborate an alternative way to characterize the asymptotic independence and asymptotic dependence. In factual terms, we investigate the relationship between tail dependence and the conditional version of a classical measure of association, namely Kendall's tau. While estimating the univariate extreme events has become a standard procedure, dealing with multivariate extreme events is a more complicated problem, and it is of general interest in many papers with particular focus on financial and insurance applications (see for example, [88] and [89]).

Some useful background is now provided for a reader that is less familiar with the justifications we made. Dependence or association is fully characterized by the copula due to the Sklar's Theorem (for example, see [3]), and for a bivariate random vector, (X_1, Y_1), is given by the joint distribution function of $\Big(F_1\big(X_1\big), F_2\big(Y_1\big)\Big)$, whenever the marginal distribution functions are continuous. Since (4.4) concerns the upper tail dependence, it is natural to study the survival copula

$$C(x, y) := \Pr\big(1 - F_1(X_1) \leq x, 1 - F_2(Y_1) \leq y\big). \tag{4.5}$$

Although the dependence is fully described by its copula or survival copula, it is sometimes difficult to explain the chosen model. The problem becomes more acute when extreme events are concerned. Instead of fully exploring the associated copula, a practical methodology is to focus on some measures of association that provide sufficient information to understand which model would be more appropriate. There are various measures of association proposed in the literature, and one of them is the Kendall's tau which is closely related to tail dependence and is defined as

$$\tau = \Pr\big((U_1 - U_2)(V_1 - V_2) > 0\big) - \Pr\big((U_1 - U_2)(V_1 - V_2) < 0\big),$$

where $U_i = 1 - F_1(X_i)$ and $V_i = 1 - F_2(Y_i)$ for $i = 1, 2$. It is well-known that this measure is scale-invariant, and therefore robust, marginal-free whenever the marginal distributions are continuous, and is based on the concept of concordance and discordance (for more details, see [4]). As a result of such appealing properties, Kendall's tau has been found useful in various fields, such as risk management (see [90]). However, if one is interested in evaluating the strength of dependence in the lower tail, when concomitant extreme events are plausible, then the conditional Kendall's tau is more sound, which is defined as follows:

$$\begin{aligned} \tau(u) \quad = \quad & \Pr\big((U_1-U_2)(V_1-V_2)>0|U_1,U_2,V_1,V_2\le u\big) \\ & -\Pr\big((U_1-U_2)(V_1-V_2)<0|U_1,U_2,V_1,V_2\le u\big). \end{aligned} \tag{4.6}$$

Study of conditional Kendall's tau for a fixed level u is relatively known in the literature (see [91] and [92]). However, it remains unknown whether there exists some relationship between the limit of this conditional measure and asymptotic dependence, and how to estimate the limit.

In the next section, we shall show that $\theta^\tau := \lim_{u\to 0}\tau(u)$ are positive for a subclass of asymptotic dependence and non-positive for a subclass of asymptotic independence. We found that all well-known examples indicate a positive limit for the case of asymptotic dependence. It is known that testing for asymptotic dependence against asymptotic independence becomes quite challenging when η is close to one. Since $\theta^\tau > 0$ may be a bit far away from zero in case of asymptotic dependence, testing for $\theta^\tau = \theta_0$ against $\theta^\tau \le 0$ becomes much easier in the case of asymptotic dependence, where θ_0 is a given positive value. That is, intervals of θ^τ are useful in distinguishing asymptotic dependence from asymptotic independence. On the other hand, when the data has the asymptotic independence property, a test based on θ^τ is less efficient than a test based on η since θ^τ may be zero, while the true value of η, say η_0, is less than one, which can be used to effectively test for $\eta = \eta_0$ against $\eta = 1$. In other words, an interval of η is quite informative when the data has the asymptotic independence property. Given the above arguments, we argue that interval estimation of $\theta^\tau + \eta$ can be effective in distinguishing between asymptotic dependence and asymptotic independence since $\theta^{\tau} + \eta$ is larger than one in case of asymptotic dependence and less than one in case of asymptotic independence. Some nonparametric estimators for the limit of this conditional measure and its asymptotic distribution are derived in Section 4.2. A set of examples, a simulation study and some empirical analyses are given in Sections 4.3, 4.4 and 4.5, respectively. Finally, all technical proofs are relegated in Section 4.6.

4.2 Main Results

A summary of our initial assumptions needed to develop our results is that $\{(X_i, Y_i)\}_{i=1}^n$ are independent and identically distributed with distribution function F, continuous marginal distribution functions F_1 and F_2, and survival copula C as defined in (4.5).

4.2.1 Conditional Kendall's tau

First, we derive the limits of the conditional Kendall's tau defined in (4.6) by assuming the following multivariate regular variation, which has been found useful in characterizing tail behavior of a random vector. Some recent references on multivariate regular variation are [93], [94, 95], and [96]).

We define $h(x, y) = \frac{\partial^2}{\partial x \partial y} H(x, y)$, $H_1(x, y) = \frac{\partial}{\partial x} H(x, y)$, $H_2(x, y) = \frac{\partial}{\partial y} H(x, y)$, $H_{11}(x, y) = \frac{\partial}{\partial x} H_1(x, y)$ and $H_{22}(x, y) = \frac{\partial}{\partial y} H_2(x, y)$, whenever the partial derivatives exist.

Assumption 4.2.1. *There exist a constant* $\delta > 0$ *and a function* $H(x, y)$ *such that* $C(u, u) > 0$ *for all* $u \in (0, \delta)$ *and*

$$H(x, y) := \lim_{u \downarrow 0} \frac{C(ux, uy)}{C(u, u)}$$

for all $(x, y) \in \mathcal{D} := [0, 1]^2$. *In addition,* $H(\cdot)$ *is continuous on* $\{(x, y) : xy = 0\}$.

Theorem 4.2.1. *Under Assumption 4.2.1, we have*

$$\theta^\tau = 4\int_0^1\int_0^1 H(x,y)dH(x,y) - 1. \tag{4.7}$$

Remark 4.2.1. *The above limit in (4.7) is indeed a proper Kendall's tau, which measures the association between two random variables with joint distribution function given by H. If the latter distribution function has continuous marginals, then one may extract the associated copula, C_H, as a result of Sklar's Theorem, and therefore (4.7) can be rewritten as follows:*

$$\theta^\tau = 4\int_{\mathcal{D}} C_H(x,y)\; dC_H(x,y) - 1 = 1 - 4\int_{\mathcal{D}} \frac{\partial}{\partial x} C_H(x,y) \frac{\partial}{\partial y} C_H(x,y)\; dx\; dy$$

(see Theorems 5.1.1. and 5.1.5 of [4]). Finally, if H admits partial derivatives, then one may show that

$$\theta^\tau = 1 - 4\int_{\mathcal{D}} H_1(x,y)H_2(x,y)\; dx\; dy.$$

Note that Assumption 4.2.1 implies that the next weak convergence

$$\mu_u(\cdot) := \Pr\big((U/u, V/u) \in \cdot | U, V \le u\big) \xrightarrow{w} \mu(\cdot) \tag{4.8}$$

holds on $\mathcal{D}$ as $u \to 0$, where the (probability) measure μ is given by $\mu\big([0,x]\times[0,y]\big) := H(x,y)$. In addition, $H(x,y)$ is a homogeneous function (see [97] and [98]). Next, we show that the limit of the conditional Kendall's tau is positive for a subclass of asymptotic dependence and non-positive for a subclass of asymptotic independence as follows:

Assumption 4.2.2. *There exist a constant $c \in [0,1]$ and an $\eta \in (0,1]$ such that*

$$H(ax,ay) = a^{1/\eta}H(x,y) \quad \text{and} \quad \lim_{u\downarrow 0} u^{-1}C(u,u) = c \in [0,1]$$

for all $a > 0$ and $(x,y) \in \mathcal{D}$.

Assumption 4.2.3. $H(x,y) = \sum_{i=1}^{m} c_i x^{\alpha_i} y^{\beta_i}$ *for some positive* $c_i's$ *and some nonnegative* $\alpha_i's, \beta_i's$ *with* $\alpha_i + \beta_i = 1/\eta$ *for* $i = 1, \cdots, m$ *and* $\sum_{i=1}^{m} c_i = 1$.

We first investigate the properties of a bivariate distribution function $H : [0,1]\times[0,1] \to [0,1]$, for which all first and second partial derivatives exist, satisfying the homogeneity property

$$H(tu, tv) = tH(u,v) \text{ for all } t > 0 \text{ and } (u,v) \in \mathcal{D}. \tag{4.9}$$

Let $\mathcal{H}$ be the collection of all such H. Define $\mathcal{F}(\xi)$, for $0 < \xi < 1$, the set of all pairs (f_X, f_Y) of density functions on $(0,1)$ such that both f_X and f_Y are non-increasing (hence almost everywhere differentiable) and

$$\int_0^x f_X(u)\, du \geq x, \quad \int_0^y f_Y(v)\, dv \geq y, \quad \lim_{x\to 1} f_X(x) = \xi, \quad \lim_{y \to 1} f_Y(y) = 1 - \xi.$$

We also define $\mathcal{F} = \bigcup_{0<\xi<1} \mathcal{F}(\xi)$. The next proposition shows that there is a one-to-one correspondence between $\mathcal{H}$ and $\mathcal{F}$.

Proposition 4.2.1. *i) Let* $H \in \mathcal{H}$ *and define* $f_X(x) = H_1(x,1)$, $f_Y(y) = H_2(1,y)$, $h(x,y) = H_{12}(x,y)$. *Then,* $(f_X, f_Y) \in \mathcal{F}$ *and for all* $(x,y), (u,v) \in \mathcal{D}$ *we have*

$$\begin{aligned} h(x,y) &= -\tfrac{x}{y^2} f_X'\left(\tfrac{x}{y}\right) I_{x<y} - \tfrac{y}{x^2} f_Y'\left(\tfrac{y}{x}\right) I_{y<x} \\ H(u,v) &= vF_X\left(\tfrac{u}{v}\right) I_{u<v} + uF_Y\left(\tfrac{v}{u}\right) I_{v\leq u}. \end{aligned} \tag{4.10}$$

ii) Let $(f_X, f_Y) \in \mathcal{F}$. *Define* $h(x,y)$ *by (4.10) and* $H(u,v) = \int_0^u \int_0^v h(x,y)\, dy\, dx$. *Then,* H *is a bivariate distribution function with marginal densities* f_X *and* f_Y *and satisfies (4.9).*

Proposition 4.2.1 allows us to identify a sharp lower bound for θ^τ and is given as Theorem 4.2.2.

Theorem 4.2.2. *Under Assumptions 4.2.1 and 4.2.2, if $\eta = 1$, $c > 0$, and $\frac{\partial^2}{\partial x^i \partial y^j} H(x, y)$ exists for all $(x, y) \in \mathcal{D}$, $i, j = 0, 1, 2$ and $i + j = 2$, then $\theta^\tau \geq -\frac{1}{2} + \frac{1}{\log(2/c)}$. Therefore, $\theta^\tau > 0$ if $c > 2e^{-2}$.*

Theorem 4.2.3. *If Assumption 4.2.3 holds, then $\lim\limits_{u \downarrow 0} \tau(u) \leq 0$.*

Remark 4.2.2. *It is clear that asymptotic dependence holds under Assumptions 4.2.1 and 4.2.2 with $\eta = 1$ and $c > 0$. Although Theorem 4.2.2 gives a lower bound on c to ensure a positive limit for the conditional Kendall's tau, a study of some common copulas indicates the limit is positive for all $c \in (0, 1]$ in the case of asymptotic dependence (see Section 3 below). Therefore it remains interesting to find a subclass of $\mathcal{H}$, which includes all $c \in (0, 1]$ and gives a positive limit.*

Remark 4.2.3. *Note that $H(x, y) \leq \min\{x, y\}/c$ for all $(x, y) \in \mathcal{D}$ due to the fact that $C(ux, uy) \leq u \min\{x, y\}$, where c is defined in Assumption 4.2.2. If Assumption 4.2.3 holds with $\eta = 1$ and $c > 0$ given in Assumption 4.2.2, then $\sum_{i=1}^m c_i (y/x)^{\beta_i} \leq c^{-1}$ and $\sum_{i=1}^m c_i (x/y)^{\alpha_i} \leq c^{-1}$ for all $(x, y) \in \mathcal{D}$, which can not be true by taking either x or y small enough. Therefore, Assumption 4.2.3 does imply the asymptotic independence. Whenever the limiting function H is not absolutely continuous, Example 4.3.4 with $\alpha = \beta \in (0, 1)$ from Section 3 illustrates that $\lim_{u \downarrow 0} \tau(u)$ may be positive for the case of asymptotic independence. Although we conjecture that $\lim_{u \downarrow 0} \tau(u) \leq 0$ for the case of asymptotic independence when $H(x, y)$ is absolutely continuous with second order partial derivatives, Theorem 4.2.3 only shows that this is true for a subclass of asymptotic independence, as defined in Assumption 4.2.3.*

Remark 4.2.4. *Example 4.3.4 with $\alpha = \beta \in (0, 1)$ from Section 3 has some positive mass along the diagonal line $y = x$, which gives a positive value for $\lim_{u \downarrow 0} \tau(u)$ for this situation of asymptotic independence. However, if one slightly modifies the definition of Kendall's* *tau as follows*

$$\begin{aligned}\tilde{\tau}(u) &= \Pr\big((U_1-U_2)(V_1-V_2)>0, U_1\neq V_1, U_2\neq V_2 | U_1,U_2,V_1,V_2\leq u\big)\\ &\quad - \Pr\big((U_1-U_2)(V_1-V_2)<0, U_1\neq V_1, U_2\neq V_2 | U_1,U_2,V_1,V_2\leq u\big),\end{aligned}$$

then it can be shown that $\theta^\tau \leq 0$ for this example. Obviously, this modification does not affects the limit of the original definition of conditional Kendall's tau when H is continuous.

4.2.2 Estimation procedure

Theorems 4.2.2 and 4.2.3 show that the limit of conditional Kendall's tau may give a good insight on whether the underlying distribution is asymptotically independent or asymptotically dependent. Hence, estimating the limit is useful in applying Extreme Value Theory to predict extreme co-movements in financial markets.

Define $\hat{F}_1(x) = \frac{1}{n+1}\sum_{i=1}^n I(X_i \leq x)$, $\hat{F}_2(y) = \frac{1}{n+1}\sum_{i=1}^n I(Y_i \leq y)$, $\hat{U}_i = 1-\hat{F}_1(X_i)$, $\hat{V}_i = 1-\hat{F}_2(Y_i)$, and put $\theta^\tau = \lim_{u\downarrow 0}\tau(u)$. Then, we propose to estimate θ^τ by

$$\hat{\theta}^\tau(k) = \frac{\sum_{1\leq i<j\leq n} sgn\Big\{\big(\hat{U}_i-\hat{U}_j\big)\big(\hat{V}_i-\hat{V}_j\big)\Big\} I\Big(\max\big(\hat{U}_i,\hat{U}_j,\hat{V}_i,\hat{V}_j\big)\leq k/n\Big)}{\sum_{1\leq i<j\leq n} I\Big(\max\big(\hat{U}_i,\hat{U}_j,\hat{V}_i,\hat{V}_j\big)\leq k/n\Big)},$$

where $k = k(n)\to\infty$ and $k/n\to 0$ as $n\to\infty$. The following theorem shows the consistency of the proposed estimator.

Theorem 4.2.4. *Under Assumption 4.2.1, $k=k(n)\to\infty$, $k/n\to 0$ and $nC\left(\frac{k}{n},\frac{k}{n}\right)\to\infty$ as $n\to\infty$, we have $\hat{\theta}^\tau(k)\xrightarrow{p}\theta^\tau$ as $n\to\infty$.*

As usual in Extreme Value Theory, if one is interested in deriving the asymptotic limit of $\hat{\theta}^\tau(k)$, a rate of convergence in (4.5) is needed, which controls the asymptotic bias of the studied estimator. Here, we employ the following second order condition.

Assumption 4.2.4. *There exist a regular variation $A(u)\to 0$ with index $\tilde{\rho}\geq 0$, i.e.*

$\lim_{u\to 0} A(ux)/A(u) = x^{\tilde{\rho}}$ *for* $x > 0$, *functions* $Q(x, y)$ *and* $q(x, y)$ *such that*

$$\lim_{u\downarrow 0} \frac{\frac{C(ux,uy)}{C(u,u)} - H(x,y)}{A(u)} = Q(x,y) \quad \textit{and} \quad \lim_{u\downarrow 0} \frac{\frac{u^2 C_{12}(ux,uy)}{C(u,u)} - H_{12}(x,y)}{A(u)} = q(x,y) \tag{4.11}$$

for all $(x, y) \in \mathcal{D}$ *and uniformly on* $\{(x, y) : x^2 + y^2 = 1\}$, *where* H_{12} *and* C_{12} *are the densities of* H *and* C, *respectively.*

Remark 4.2.5. *The second condition in (4.11) implies the first one when some mild integrability conditions are satisfied.*

Theorem 4.2.5. *Under Assumption 4.2.4,* $\lim_{u\downarrow 0} u^{-1}C(u, u) = c \in [0, 1]$,

$$k = k(n) \to \infty,\ nC\left(\frac{k}{n}, \frac{k}{n}\right) \to \infty \ \textit{ and } \sqrt{nC\left(\frac{k}{n}, \frac{k}{n}\right)} A\left(\frac{k}{n}\right) \to \lambda \in (-\infty, \infty)$$

as $n \to \infty$, *we have*

$$\sqrt{nC\left(\frac{k}{n}, \frac{k}{n}\right)} \big\{\hat{\theta}^{\tau}(k) - \theta^{\tau}\big\} \xrightarrow{d} N\big(\lambda b_{\tau}, \sigma_{\tau}^2\big) \tag{4.12}$$

as $n \to \infty$, *where*

$$b_{\tau} = 4\int_0^1 \int_0^1 Q(s,t) H_{12}(s,t)\, dtds + 4\int_0^1 \int_0^1 H(s,t) q(s,t)\, dtds,$$

$$\sigma_{\tau}^2 = 4\{\sigma_1^2 - (\theta^{\tau})^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_2\sigma_3 c\}, \tag{4.13}$$

with

$$
\begin{cases}
\sigma_1^2 & = 16\int_0^1\int_0^1 H^2(x,y)\,dH(x,y) - 16\int_0^1\int_0^1 H(x,1)H(x,y)\,dH(x,y) \\
& \quad -16\int_0^1\int_0^1 H(1,y)H(x,y)\,dH(x,y) + 8\int_0^1\int_0^1 H(x,y)\,dH(x,y) \\
& \quad +8\int_0^1\int_0^1 H(x,1)H(1,y)\,dH(x,y) - \frac{1}{3} \\
\sigma_2 & = \sqrt{c}(2\int_0^1 H(1,t)H_{12}(1,t)\,dy - H_1(1,1)) \\
\sigma_3 & = \sqrt{c}(2\int_0^1 H(s,1)H_{12}(s,1)\,ds - H_2(1,1)).
\end{cases}
\tag{4.14}
$$

Remark 4.2.6. *When $C(u,u) = d_1 u^{1/\eta}$ and $A(u) = d_2 u^{\tilde{\rho}}$, a theoretical optimal k for $\hat{\theta}^\tau(k)$ can be chosen to minimize the asymptotic mean squared error $b_\tau^2 A^2\left(\frac{k}{n}\right) + \frac{\sigma_\tau^2}{nC\left(\frac{k}{n},\frac{k}{n}\right)}$, which gives the optimal choice of k as*

$$k_0^\tau = \left(\frac{\sigma_\tau^2}{2\eta b_\tau^2 d_2^2 d_1 \tilde{\rho}}\right)^{1/(2\tilde{\rho}+1/\eta)} n^{(1/\eta-1+2\tilde{\rho})/(1/\eta+2\tilde{\rho})}.$$

Remark 4.2.7. *A consistent estimator for σ_τ^2 can be obtained by replacing c, $H(x,y)$ and $H_{12}(x,y)$ in (4.13) and (4.14) by*

$$\hat{c} = \frac{1}{m}\sum_{i=1}^n I\left(1-\hat{F}_1(X_i)\frac{m}{n}, 1-\hat{F}_2(Y_i) \le \frac{m}{n}\right),$$

$$\hat{H}(x,y) = \frac{1}{m\hat{c}}\sum_{i=1}^n I\left(1-\hat{F}_1(X_i) \le \frac{m}{n}x, 1-\hat{F}_1(Y_i) \le \frac{m}{n}y\right),$$

$$
\begin{aligned}
\hat{H}_{12}(x,y) \quad = \quad & \sum_{i=1}^n \frac{I\left(1-\hat{F}_1(X_i)\le\frac{m}{n}x, 1-\hat{F}_1(Y_i)\le\frac{m}{n}y\right)}{m\hat{c}}\, G\left(\frac{\frac{n}{m}\left(1-\hat{F}_1(X_i)\right)-x}{q}\right) \\
& \times G\left(\frac{\frac{n}{m}\left(1-\hat{F}_2(Y_i)\right)-y}{q}\right),
\end{aligned}
$$

respectively, where $m = m(n) \to \infty$, $m/n \to 0$ as $n \to \infty$, G is a smooth distribution function and $q = q(n) > 0$ is the bandwidth satisfying that $q \to 0$ and $qm \to \infty$ as $n \to \infty$. One can also use the corresponding estimators in [86].

In the simulation study, we employ the bootstrap method to estimate the asymptotic variance. Theoretical justification of the proposed bootstrap method can be shown in a *similar way to [99].*

Remark 4.2.8. *The usual approach to construct confidence intervals for θ is to choose $k = o(k_0^\tau)$ so that the asymptotic bias is negligible, where k_0^τ is the theoretical optimal choice given in Remark 4.2.6. Motivated by the choice of sample fraction for the Hill estimator in terms of coverage probability in [100], we propose to choose $k = O\left(n^{(1/\eta-1+\tilde{\rho})/(1/\eta+\tilde{\rho})}\right)$ for interval estimation of θ^τ based on the asymptotic limits of $\hat{\theta}^\tau(k)$.*

Remark 4.2.9. *As argued in the introduction, when the datum is asymptotically independent, θ^τ may be zero, hence the interval may not be effective in distinguishing the asymptotic independence from the asymptotic dependence. In this case, one may use the quantity $\theta^\tau+\eta$. For estimating $\theta^\tau+\eta$, one can easily combine $\hat{\theta}^\tau$ with the estimator $\hat{\eta}$ for η proposed in [86], and the asymptotic distribution of $\hat{\theta}^\tau + \hat{\eta}$ can be derived by using expansions as given in the proof of Theorem 4.2.5 and those in [86], but we skip these derivations. For constructing an interval for $\theta^\tau+\eta$ based on the normal approximation of $\hat{\theta}^\tau+\hat{\eta}$, we simply employ the bootstrap method as we do in Section 4.5.*

4.3 Examples

This section shows that some well-known copulas satisfy the conditions from Theorems 4.2.2 and 4.2.3 for which the limit of the conditional Kendall's tau is also derived. If C^* is a copula with corresponding survival copula C defined in (4.5), then $C(u,v) = C^*(1-u, 1-v)+u+v-1$ for all $(u,v) \in \mathcal{D}$.

Example 4.3.1. *Consider the Gumbel copula $C^*(u,v) = \exp\left\{-\left((-\log u)^\alpha+(-\log v)^\alpha\right)^{1/\alpha}\right\}$ where $\alpha \in (1,\infty)$. Then, Assumption 4.2.2 holds with $\eta = 1$, $c = 2-2^{1/\alpha}$ and $cH(x,y) = x+y-(x^\alpha+y^\alpha)^{1/\alpha}$. Figure 4.1 below plots the values of θ^τ against different α, which shows that the limit is positive.It is easy to show that $H_1(x,1)$ increases in α for $x \in (0,1]$ and so is the limit of conditional Kendall's tau. By $\lim_{\alpha\to1} H_1(x,1) = \frac{\ln(1+x)}{2\ln 2}$ and $\ln(1+x) \le$*

$\frac{x}{\sqrt{1+x}}$ *for* $x > 0$, *we have*

$$\begin{aligned}
\lim_{u\to 0}\theta^{\tau} &\geq 4\int_0^1 \lim_{\alpha\to 1} xH_1^2(x,1)dx - 1 \\
&= 1 - 4\int_0^1 \lim_{\alpha\to 1} H_1(x,1)H_1(1,x)dx \\
&= 1 - \int_0^1 \frac{\ln(1+x)\ln(1+x^{-1})}{(\ln 2)^2}dx \\
&\geq 1 - \int_0^1 \frac{\frac{x}{\sqrt{1+x}}\frac{x^{-1}}{\sqrt{1+x^{-1}}}}{(\ln 2)^2}dx \\
&\geq 1 - \frac{1}{\sqrt{6}(\ln 2)^2} \\
&> 0.
\end{aligned}$$

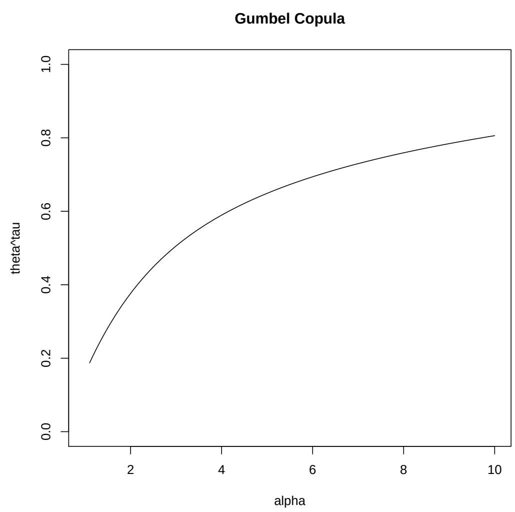


Figure 4.1: The limit of conditional Kendall's tau is plotted against parameter for Gumbel copula from Example 4.3.1.

Example 4.3.2. *Consider the* t *copula*

$$C^*(u,v) = \int_{-\infty}^{t_\nu^-(u)}\int_{-\infty}^{t_\nu^-(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}}\left\{1+\frac{x^2-2\rho xy+y^2}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} dxdy,$$

where $|\rho| < 1$, $\nu > 0$ *and* t_ν *denotes the distribution function of a* t *distribution with* ν *degrees of freedom. Let* (U_1^*, V_1^*) *be a bivariate random vector with distribution* C^*. *Since*

$t_\nu^-(1-s) \sim ds^{-1/\nu}$ for some constant $d>0$ as $s \to 0$, we have

$$\begin{aligned}
&\lim_{s\to 0}\frac{1-C^*(1-su,1-sv)}{s}\\
=\;& u\lim_{s\to 0}\Pr\left(V_1^* \le 1-sv | U_1^* = 1-su\right) + v\lim_{s\to 0}\Pr\left(U_1^* \le 1-su|V_1^*=1-sv\right)\\
=\;& u\lim_{s\to 0}\Pr\left(t_\nu^-(V_1^*) \le t_\nu^-(1-sv)|t_\nu^-(U_1^*) = t_\nu^-(1-su)\right)\\
&+v\lim_{s\to 0}\Pr\left(t_\nu^-(U_1^*) \le t_\nu^-(1-su)|t_\nu^-(V_1^*) = t_\nu^-(1-sv)\right)\\
=\;& u\lim_{s\to 0} t_{\nu+1}\left(\frac{t_\nu^-(1-sv)-\rho t_\nu^-(1-su)}{\sqrt{1-\rho^2}}\left(\frac{\nu+1}{\nu+\left(t_\nu^-(1-su)\right)^2}\right)^{1/2}\right)\\
&+v\lim_{s\to 0} t_{\nu+1}\left(\frac{t_\nu^-(1-su)-\rho t_\nu^-(1-sv)}{\sqrt{1-\rho^2}}\left(\frac{\nu+1}{\nu+\left(t_\nu^-(1-sv)\right)^2}\right)^{1/2}\right)\\
=\;& ut_{\nu+1}\left(\frac{\left((v/u)^{-1/\nu}-\rho\right)\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\right)+vt_{\nu+1}\left(\frac{\left((u/v)^{-1/\nu}-1\right)\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\right).
\end{aligned}$$

Consequently, Assumption 4.2.2 holds with $\eta=1$, $c = 2-2t_{\nu+1}\left(\sqrt{\frac{(1-\rho)(\nu+1)}{1+\rho}}\right)$ and

$$\begin{aligned}
cH(x,y) &= x\left\{1-t_{\nu+1}\left(\frac{\left((y/x)^{-1/\nu}-\rho\right)\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\right)\right\}\\
&\quad +y\left\{1-t_{\nu+1}\left(\frac{\left((x/y)^{-1/\nu}-\rho\right)\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\right)\right\}.
\end{aligned}$$

Figure 4.2 below plots the values of θ^τ against various ρ and ν, which shows that the limit is indeed positive.

Example 4.3.3. *Consider the elliptical copula $Z \overset{d}{=} GAU$, where $G>0$ is a random variable with a survival function, $\bar{G}(\cdot)$, that satisfies $\bar{G}(tx)/\bar{G}(t) \sim x^{-\alpha}$ as $t\to\infty$ for all $x>0$, A is a deterministic 2×2 matrix with $AA^T = \begin{pmatrix} 1 & \rho \\ \rho & 1\end{pmatrix}$ with $|\rho|<1$, U is uniformly*

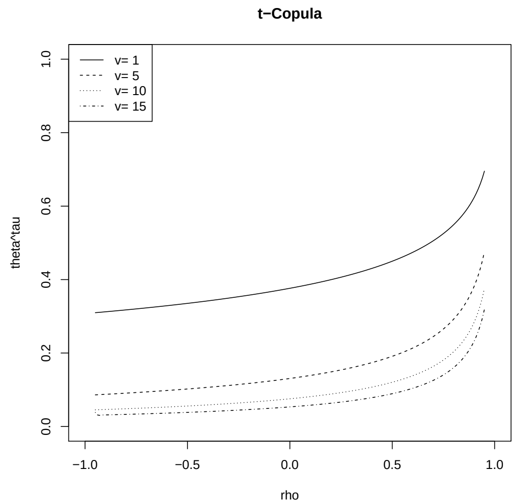


Figure 4.2: The limit of conditional Kendall's tau is plotted against parameters for t copula from Example 4.3.2.

distributed on $\{z \in \Re^2 : z^T z = 1\}$ *and independent of* G*. Put*

$$\lambda(x,y) = \frac{x \int_{g\left((x/y)^{1/\alpha}\right)}^{\pi/2} (\cos\phi)^\alpha \, d\phi + y \int_{g\left((x/y)^{-1/\alpha}\right)}^{\pi/2} (\cos\phi)^\alpha \, d\phi}{\int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha \, d\phi}, \tag{4.15}$$

where $g(t) = \arctan\left((t-\rho)/\sqrt{1-\rho^2}\right)$*. Then it follows from [101] that Assumption 4.2.2 holds with* $\eta = 1$*,* $c = \lambda(1,1)$ *and* $H(x,y) = \lambda(x,y)/\lambda(1,1)$*. Figure 4.3 below plots the values of* θ^τ *against various* ρ *and* α*, which shows that the limit is indeed positive. A rigorous verification goes as follows.*

It is easy to check that

$$\begin{cases} g(t) + g(t^{-1}) = \arccos\rho, \; t > 0 \\ cos(g(t)) = (1 + (\frac{t-\rho}{\sqrt{1-\rho^2}})^2)^{-\frac{1}{2}} = \sqrt{(1-\rho^2)^{\frac{1}{2}} g'(t)} = t^{-1}\cos(g(t^{-1})), t > 0 \end{cases} \tag{4.16}$$

where g' *is the derivative of* g *with respect to* t*. Taking the partial derivatives of* $\lambda(x,y)$*, by*

(4.16), it follows that

$$\begin{cases} \frac{\partial}{\partial x}\lambda(x,y) = \left\{ \int_{g((\frac{x}{y})^{\frac{1}{\alpha}})}^{\pi/2} (\cos\phi)^\alpha d\phi \right\} \times \left\{ \int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha d\phi \right\}^{-1}, \\ \frac{\partial}{\partial y}\lambda(x,y) = \left\{ \int_{g((\frac{x}{y})^{-\frac{1}{\alpha}})}^{\pi/2} (\cos\phi)^\alpha d\phi \right\} \times \left\{ \int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha d\phi \right\}^{-1}. \end{cases} \tag{4.17}$$

Define $D(t,\rho) = \sqrt{1-\rho^2}\int_{g(t)}^{\pi/2}(\cos\phi)^\alpha d\phi, t > 0$*, then* $D(t,\rho)$ *is strictly decreasing in* t *and has the following properties,*

$$\begin{cases} D'(t,\rho) = \frac{d}{dt}D(t,\rho) = -(\cos(g(t)))^{\alpha+2} \\ D(t,\rho) = \int_t^\infty (\cos(g(s)))^{\alpha+2}ds = \int_0^{t^{-1}} s^\alpha(\cos(g(s)))^{\alpha+2}ds \\ D(t,\rho) < D(0^+,\rho) = \lim_{t\to 0^+} D(t,\rho) < \infty \\ D(t,\rho) > D(\infty,\rho) = \lim_{t\to\infty} D(t,\rho) = 0. \end{cases} \tag{4.18}$$

Further

$$H_1(x,1) = H_2(1,x) = \frac{D(x^{\frac{1}{\alpha}},\rho)}{2D(1,\rho)}. \tag{4.19}$$

Since the elliptical copula is symmetric, we also have $H_1(1,x) = H_2(x,1)$*. Put them into (4.7) we have*

$$\begin{aligned} \lim_{u\downarrow 0}\tau(u) &= 4\int_0^1\int_0^1 H(x,y)dH(x,y) - 1 \\ &= 2\int_0^1 xH_1^2(x,1)dx + 2\int_0^1 yH_2^2(1,y)dy - 1 \\ &= 4\int_0^1 xH_1^2(x,1)dx - 1 \\ &= 4(\tfrac{1}{2} - \int_0^1 H_1(x,1)H_2(x,1)dx) - 1 \\ &= 1 - 4\int_0^1 H_1(x,1)H_1(1,x)dx \\ &= 1 - \int_0^1 \frac{D(x^{\frac{1}{\alpha}},\rho)D(x^{-\frac{1}{\alpha}},\rho)}{D^2(1,\rho)}dx. \end{aligned} \tag{4.20}$$

Hence, to show the limit is positive, it is equivalent to show that

$$\int_0^1 D(x^{\frac{1}{\alpha}},\rho)D(x^{-\frac{1}{\alpha}},\rho)dx < D^2(1,\rho).$$

which is sufficiently implied by

$$D(t,\rho)D(t^{-1},\rho) < D^2(1,\rho), \quad 0<t<1. \tag{4.21}$$

By (4.18), for $0<t<1$ *we have*

$$\begin{aligned} D(t,\rho)D(t^{-1},\rho) &= \Big(D(1,\rho)+\textstyle\int_t^1(\cos(g(s)))^{\alpha+2}ds\Big) \\ &\quad \times\Big(D(1,\rho)-\textstyle\int_1^{t^{-1}}(\cos(g(s)))^{\alpha+2}ds\Big) \\ &= D^2(1,\rho)+\Big(\textstyle\int_t^1(\cos(g(s)))^{\alpha+2}ds \\ &\quad -\textstyle\int_1^{t^{-1}}(\cos(g(s)))^{\alpha+2}ds\Big)D(1,\rho) \\ &\quad -\textstyle\int_t^1(\cos(g(s)))^{\alpha+2}ds\int_1^{t^{-1}}(\cos(g(s)))^{\alpha+2}ds. \end{aligned} \tag{4.22}$$

Put $a=\int_t^1(\cos(g(s)))^{\alpha+2}ds$ *and* $b=\int_1^{t^{-1}}(\cos(g(s)))^{\alpha+2}ds=\int_t^1 s^\alpha(\cos(g(s)))^{\alpha+2}ds$*, let* a',b' *be the derivatives of functions* a *and* b *with respect to* t*. It follows that* $a>b>0$ *and* $a'<b'<0$*, and thus (4.21) is equivalent to*

$$D(1,\rho) < \frac{ab}{a-b} \tag{4.23}$$

and taking the derivative of the left side of (4.23), we have

$$\frac{d}{dt}\Big(\frac{ab}{a-b}\Big)=\frac{a^2b'-a'b^2}{(a-b)^2} > \frac{a^2a'-a'a^2}{(a-b)^2}=0 \tag{4.24}$$

Therefore,

$$\begin{aligned} \tfrac{ab}{a-b} &\geq \tfrac{\int_0^1(\cos(g(s)))^{\alpha+2}ds\int_1^\infty(\cos(g(s)))^{\alpha+2}ds}{\int_0^1(\cos(g(s)))^{\alpha+2}ds-\int_1^\infty(\cos(g(s)))^{\alpha+2}ds} \\ &= \tfrac{D(0,\rho)-D(1,\rho)}{D(0,\rho)-2D(1,\rho)}D(1,\rho) \\ &> D(1,\rho) \end{aligned} \tag{4.25}$$

which implies the limit of conditional Kendall's tau is positive.

Example 4.3.4. *Assume that the survival copula is given by the Marshall-Olkin copula.*

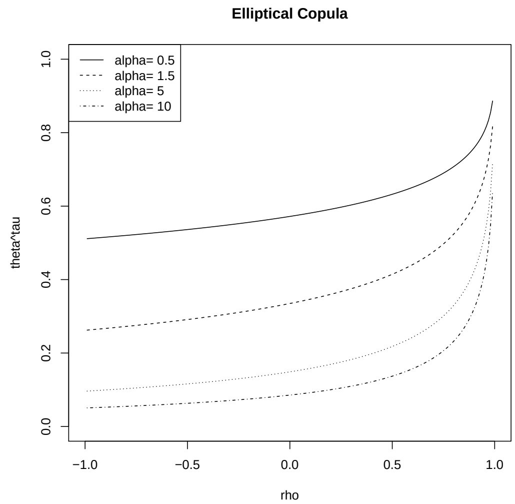


Figure 4.3: The limit of conditional Kendall's tau is plotted against parameters for elliptical copula from Example 4.3.3.

That is, we have

$$C(u,v) = \begin{cases} u^{1-\alpha}v & \textit{if} \quad u^\alpha \geq v^\beta, \\ uv^{1-\beta} & \textit{if} \quad u^\alpha < v^\beta, \end{cases} \tag{4.26}$$

where $0 < \alpha, \beta < 1$*. Simple calculations yield that Assumption 4.2.1 holds with*

$$H(x,y) = \begin{cases} xy^{1-\beta} & \textit{if} \quad \alpha > \beta, \\ x^{1-\alpha}y & \textit{if} \quad \alpha < \beta, \\ xy\big(\max\{x,y\}\big)^{-\alpha} & \textit{if} \quad \alpha = \beta. \end{cases}$$

Therefore, Assumption 4.2.3 holds with $\eta = \big(2 - \min\{\alpha,\beta\}\big)^{-1}$*,* $m = 1$*, and* $\theta^\tau = 0$ *for* $\alpha \neq \beta$*. When* $\alpha = \beta$*,* $\eta = (2-\alpha)^{-1}$*,* $H(x,y)$ *has a positive mass along the line* $y = x$ *and Assumption 4.2.3 does not hold. In this case, some straightforward computations lead to* $\Pr(U = V \leq z) = \frac{\alpha}{2-\alpha}z^{2-\alpha}$ *for* $0 \leq z \leq 1$*,* $\theta^\tau = \frac{4}{4-2\alpha} - 1 = \frac{\alpha}{2-\alpha} > 0$*, where* (U,V) *has the distribution* $C(u,v)$ *given in (4.26).*

Example 4.3.5. *Consider the bivariate normal copula*

$$C^*(u,v) = \int_{-\infty}^{\Phi^-(u)} \int_{-\infty}^{\Phi^-(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right\} dydx, \quad |\rho| < 1,$$

where Φ denotes the distribution function of the standard normal random variable. Then, it follows from Example 2.1 of [86] or Theorem 5.3 of [102] that Assumption 4.2.3 holds with $H(x,y) = (xy)^{1/(1+\rho)}$ and $\eta = (1+\rho)/2$. Thus, Assumption 4.2.3 holds with $m = 1$, and $\theta^\tau = 0$. Interestingly, a more general result can be found for the class of elliptical copulas, as defined in Example 4.3.3, where $\bar{G}(\cdot)$ satisfies $\bar{G}\big(t + a(t)x\big)/\bar{G}(t) \sim e^{-x}$ and $a(ty)/a(t) \sim y^{-\alpha}$ as $t \to \infty$ for all $x \in \Re$ and $y > 0$. In has been shown in [103] that $H(x,y) = (xy)^{1/2\eta}$ where $\eta = \big(2/(1+\rho)\big)^{(\alpha-1)/2}$. Note that the Gaussian copula is a special case of this last result and it holds with $\alpha = -1$, which confirms the earlier finding. Once again, Assumption 4.2.3 holds with $m = 1$, and $\theta^\tau = 0$.

Example 4.3.6. *Consider the Farlie-Gumbel-Morgenstern copula*

$$C^*(u,v) = uv\big\{1 + \xi(1-u)(1-v)\big\} \quad \textit{with} \quad \xi \in [-1,1].$$

Simple computations yield that Assumption 4.2.1 holds with

$$H(x,y) = \begin{cases} xy & \textit{if } \xi \in (-1,1], \\ \frac{xy(x+y)}{2} & \textit{if } \xi = -1. \end{cases}$$

Hence, Assumption 4.2.3 holds with $(\eta, m) = (1/2, 1)$ for $\xi \in (-1,1]$ and $(\eta, m) = (1/3, 2)$ for $\xi = -1$. Further, $\theta^\tau = 0$ for $\xi \in (-1,1]$, and $\theta^\tau = -\frac{1}{18}$ for $\xi = -1$.

4.4 Simulation study

In this section, we examine the finite sample behavior of the proposed estimator $\hat{\theta}^\tau(k)$ for estimating the limit of conditional Kendall's tau by drawing $1,000$ random samples with size $n = 1000$ from Examples 4.3.2, 4.3.5 and 4.3.6 given in Section 4.3. For estimating the asymptotic variance of $\hat{\theta}^{\tau}(k)$ we simply employ the bootstrap method with $1,000$ resamples. Based on these random samples, we have estimators $\hat{\theta}_{\tau}^{(i)}(k)$ and the corresponding bootstrap variance estimator $\sigma^{(i)}(k)$ for $i = 1, \cdots, 1000$. In Figures 4.4,4.5,4.6 and 4.7, we plot the estimator $\frac{1}{1000}\sum_{i=1}^{1000}\hat{\theta}_{\tau}^{(i)}(k)$, the bias $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{\tau}^{(i)}(k) - \theta^{\tau})$, the mean squared error $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{\tau}^{(i)}(k)-\theta^{\tau})^2$ and the ratio of asymptotic variance to its bootstrap estimator

$$\sum_{i=1}^{1000}\left(\hat{\theta}_{\tau}^{(i)}(k) - \frac{1}{1000}\sum_{j=1}^{1000}\hat{\theta}_{\tau}^{(j)}(k)\right)^2 / \sum_{i=1}^{1000}\sigma^{(i)}(k)$$

against $k = 21, \cdots, 300$. These figures show that the estimator and its bootstrap variance estimator work well for k around 150. Without doubt, more research on choosing the tuning parameter k in estimating θ^{τ}, $\theta^{\tau} + \eta$, and corresponding bias reduced estimators is needed in the near future.

4.5 Real data analysis

In this section, we analyze the tail dependence of the following three data sets by estimating $\eta, \theta^{\tau}, \theta^{\tau}+\eta$ by $\hat{\eta}(k), \hat{\theta}^{\tau}(k), \hat{\theta}^{\tau}(k)+\hat{\eta}(k)$, respectively, where $\hat{\eta}(k)$ is the Hill estimator based on the largest k order statistics of

$$\left\{T_i = \min\{\frac{n+1}{n+1-R_i^X}, \frac{n+1}{n+1-R_i^Y}\}\right\}_{i=1}^{n}$$

with R_i^X being the rank of X_i among $X_1, \cdots, X_n$ and R_i^Y being the rank of Y_i among $Y_1, \cdots, Y_n$. More details on $\hat{\eta}(k)$ can be found in [86]. For constructing confidence intervals for η, θ^{τ}, $\theta^{\tau} + \eta$ via corresponding estimators, we simply employ the bootstrap method with $1,000$ replications.

First, we consider the sea level and wave height measured at the Eierland station, 20 km off the Dutch coast from 1979 through 1991; see the left upper panel in Figure 4.8.

The right upper panel depicts the $\hat{\eta}(k)$ and its intervals, which may suggest asymptotic independence by looking at k near 50 as argued in [86]. However, the left lower panel may well suggest $\theta^\tau > 0$ by looking at the range of $50 < k < 100$, i.e., the data set is asymptotically dependent. The right lower panel do not claim that $\theta^\tau + \eta < 1$, i.e. asymptotic independence, even when one chooses a smaller k. Therefore, it is reasonable to assume asymptotic dependence and so it is recommended to employ the asymptotic dependent classical Extreme Value Theory to predict extreme co-movements.

Next, we consider the non-zero losses to building and content in the Danish fire insurance claims; see the left upper panel in Figure 4.9. This data set is available at *www.ma.hw.ac.uk/∼mcneil/*, which comprises 2,167 fire losses over the period 1980 to 1990. The right upper panel may prefer $\eta < 1$, i.e., asymptotic independence. However, the lower panels can neither claim asymptotic independence nor asymptotic dependence. Therefore one may claim asymptotic independence for this data set. On the other hand, given the fact that distinguishing asymptotic behavior is extremely challenging, one has to take a caution of making the claim of asymptotic independence since this claim is not confirmed by the two new measures $\hat{\theta}^\tau(k)$ and $\hat{\theta}^\tau(k) + \hat{\eta}(k)$.

Finally, we consider the log-returns of the exchange rates between Euro and US dollar and those between British pound and US dollar from January 3, 2000 until December 19, 2007; see the left upper panel in Figure 4.10. The right upper panel may well suggest $\eta < 1$, i.e., asymptotic independence. The left lower panel may prefer $\theta^\tau > 0$, i.e., asymptotic dependence. The right lower panel can neither claim asymptotic independence nor asymptotic dependence. Therefore, it remains cautious to claim the asymptotic behavior for this data set, which calls for more effective methods.

In summary, the proposed new measure of tail dependence and its combination with the coefficient of tail dependence are useful in distinguishing between asymptotic dependence and asymptotic independence, so as to ensure a sound application of multivariate Extreme Value Theory to the study of extreme co-movements in financial markets and so to predicting extreme events.

4.6 Proofs

Proof of Theorem 4.2.1. Since

$$\begin{aligned}
& \Pr\big(U_1 > U_2, V_1 > V_2 | U_1, U_2, V_1, V_2 \le u\big) \\
& + \Pr\big(U_1 > U_2, V_1 < V_2 | U_1, U_2, V_1, V_2 \le u\big) \\
= \; & \Pr\big(U_1 > U_2 | \max(U_1, U_2, V_1, V_2) \le u\big) \\
= \; & \tfrac{1}{C^2(u,u)} \textstyle\int_0^u C(t,u) \Pr(U_1 \in dt, V_1 \le u) \\
= \; & \tfrac{1}{2},
\end{aligned}$$

it follows from (4.6) that

$$\begin{aligned}
\tau(u) \; = \; & 2\Pr\big(U_1 > U_2, V_1 > V_2 | U_1, U_2, V_1, V_2 \le u\big) \\
& -2\Pr\big(U_1 > U_2, V_1 < V_2 | U_1, U_2, V_1, V_2 \le u\big) \\
= \; & 4\Pr\big(U_1 > U_2, V_1 > V_2 | U_1, U_2, V_1, V_2 \le u\big) - 1.
\end{aligned} \tag{4.27}$$

Next, we define the following probability measure

$$\nu_u(\cdot) := \Pr\Big(\big(U_1/u, V_1/u, U_2/u, V_2/u\big) \in \cdot | U_1, U_2, V_1, V_2 \le u\Big)$$

on $\mathcal{E} := [0,1]^4$. Thus, due to equation (4.8) and the independence assumption between (U_1, V_1) and (U_2, V_2), we have that

$$\nu_u(\cdot) \overset{w}{\to} \nu(\cdot) \tag{4.28}$$

holds on $\mathcal{E}$ as $u \to 0$, where the measure ν is given by

$$\nu\big([0,x_1] \times [0,y_1] \times [0,x_2] \times [0,y_2]\big) := H(x_1,y_1)H(x_2,y_2).$$

Let $A := \{0 \le x_2 < x_1 \le 1, 0 \le y_2 < y_1 \le 1\}$. Therefore, relation (4.28) leads to

$$\begin{aligned} & \Pr\big(U_1 > U_2, V_1 > V_2 | U_1, U_2, V_1, V_2 \le u\big) = \nu_u(A) \\ \to \quad & \nu(A) = \textstyle\int_{\mathcal{D}} H(x,y)\, dH(x,y), \text{ as } u \downarrow 0 \end{aligned} \tag{4.29}$$

as long as $\nu(\partial A) = 0$, which remains to justify. Note that

$$\begin{aligned} \nu(\partial A) \quad \le \quad & \nu\big(x_1 = x_2, y_1 \ge y_2\big) + \nu\big(x_1 \ge x_2, y_1 = y_2\big) \\ + \quad & \nu\big(x_1 \ge x_2, y_1 \ge y_2,\ x_1 y_1 = 0 \text{ or } x_1 = 1 \text{ or } y_1 = 1\big) \\ + \quad & \nu\big(x_1 \ge x_2, y_1 \ge y_2,\ x_2 y_2 = 0 \text{ or } x_2 = 1 \text{ or } y_2 = 1\big). \end{aligned}$$

The first two terms are equal to zero since no mass is put by the measure ν over the lines $x_1 = x_2$ and $y_1 = y_2$ due to the independence between (U_1, V_1) and (U_2, V_2). The last two terms are also negligible and due to symmetry, it is sufficient to justify only one of them. Denote $B = \big\{x_1 y_1 = 0 \text{ or } x_1 = 1 \text{ or } y_1 = 1\big\}$ and note that

$$\nu\big(x_1 \ge x_2, y_1 \ge y_2,\ x_1 y_1 = 0 \text{ or } x_1 = 1 \text{ or } y_1 = 1\big) \le \int_B \mu\big(dx_1, dy_1\big) = 0,$$

since H continuous on $\{xy = 0\}$ (due to Assumption 4.2.1) and the fact that $\mu(x_1 = 1) = \mu(y_1 = 1) = 0$, where the measure μ is defined in (4.8). The later is true, since otherwise we find a contradiction as follows

$$\begin{aligned} \mu(x_1 \ge y_1 > 0) \quad \ge \quad & \mu\left(\bigcup_{q \in Q \bigcap (0,1]} \{x_1 = q, y_1 \le q\}\right) \\ = \quad & \sum_{q \in Q \bigcap (0,1]} \mu\big(\{x_1 = q, y_1 \le q\}\big) \\ = \quad & \mu\big(\{x_1 = 1\}\big) \sum_{q \in Q \bigcap (0,1]} q^a = \infty, \end{aligned}$$

where $a \ge 1$ is the homogeneous order of H. Therefore, (4.7) follows from equations (4.27) and (4.29). $\square$

Proof of Proposition 4.2.1. i) Clearly, $\int_0^u f_X(x)\,dx = H(u,1) \geq H(u,u) = uH(1,1) = u$. Similarly, one may get the mirror result for f_Y. Differentiating (4.9) with respect to t in the case $\eta = 1$, we have $H_1(tu,tv)+H_2(tu,tv) = H(u,v)$, and therefore, $f_X(1)+f_Y(1) = 1$ is true. Now, differentiating (4.9) with respect to u (respectively v), we have

$$H_1(tu,tv) = H_1(u,v) \quad \text{(respectively } H_2(tu,tv) = H_2(u,v)\text{)}.$$

Let us first look at the case $x < y$; by setting $v = 1$, $t = y$, $u = x/y$ in the above equation, we have

$$H_1(x,y) = H_1\left(\frac{x}{y},1\right) = f_X\left(\frac{x}{y}\right),$$

and in turn, differentiating with respect to y gives $h(x,y) = -\frac{x}{y^2}f'_X\left(\frac{x}{y}\right)$. Note that the left-hand side of the latter equation is a bivariate density function, and thus, it is non-negative. In addition, it follows that $f'_X \leq 0$. The same procedure can be applied in the case $y < x$ in order to justify (4.10).

Suppose that $u \leq v$. Now,

$$\begin{aligned}
H(u,v) &= \int_0^u dx \int_0^v dy\, h(x,y) \\
&= -\int_0^u dx \left\{ \int_0^x dy\, \frac{y}{x^2} f'_Y\left(\frac{y}{x}\right) + \int_x^v dy\, \frac{x}{y^2} f'_X\left(\frac{x}{y}\right) \right\} \\
&= -\int_0^u dx \left\{ \int_0^1 dw\, w f'_Y(w) + \int_{\frac{x}{v}}^1 dz\, f'_X(z) \right\} \\
&= -\int_0^u dx \left\{ [w f_Y(w)]_0^1 - \int_0^1 dw\, f_Y(w) + f_X(1) - f_X\left(\frac{x}{v}\right) \right\} \\
&= -\int_0^u dx \left\{ (1-\xi) - 1 + \xi - f_X\left(\frac{x}{v}\right) \right\} \\
&= \int_0^{\frac{u}{v}} dw\, v f_X(w) \\
&= vF_X\left(\frac{u}{v}\right).
\end{aligned}$$

Again, the same procedure can be applied for $u > v$, and thus part i) is justified.

ii) The function h is certainly non-negative, since f_X and f_Y are non-increasing functions. In addition, the integration procedure to derive H from h has been accomplished above. Moreover, it is elementary to check that $H(u,1) = F_X(u)$ and $H(1,v) = F_Y(v)$. Finally, part ii) is concluded due to

$$H(tu,tv) = I_{tu<tv} tv F_X\left(\frac{tu}{tv}\right) + I_{tu\geq tv} tu F_Y\left(\frac{tv}{tu}\right) = tH(u,v).$$

$\square$

Proof of Theorem 4.2.2. Since $H(tx,ty) = tH(x,y)$, by taking derivatives with respect to t at both sides, we have $xH_1(tx,ty) + yH_2(tx,ty) = H(x,y)$, i.e., $txH_1(tx,ty) + tyH_2(tx,ty) = tH(x,y) = H(tx,ty)$, which implies that

$$xH_1(x,y) + yH_2(x,y) = H(x,y) \quad \text{for all} \quad (x,y) \in \mathcal{D}. \tag{4.30}$$

By taking the derivative with respect to x in (4.30), one may show

$$xH_{11}(x,y) + yh(x,y) = 0 \quad \text{for all} \quad (x,y) \in \mathcal{D}. \tag{4.31}$$

Similarly, $yH_{22}(x,y) + xh(x,y) = 0$ holds for all $(x,y) \in \mathcal{D}$. By (4.30), we can write

$$\begin{aligned} \textstyle\int_0^1\int_0^1 H(x,y)h(x,y)\,dxdy &= \textstyle\int_0^1\int_0^1 xH_1(x,y)h(x,y)\,dxdy \\ &\quad \textstyle+ \int_0^1\int_0^1 yH_2(x,y)h(x,y)\,dxdy. \end{aligned} \tag{4.32}$$

It follows from (4.31) that

$$
\begin{aligned}
\int_0^1\int_0^1 xH_1(x,y)h(x,y)\,dxdy &= \int_0^1\int_0^y xH_1(x,y)h(x,y)\,dxdy && (4.33)\\
&\quad+\int_0^1\int_0^x xH_1(x,y)h(x,y)\,dydx\\
&= \int_0^1\int_0^1 xyH_1(xy,y)h(xy,y)y\,dxdy && (4.34)\\
&\quad+\int_0^1\int_0^1 xH_1(x,xy)h(x,xy)x\,dydx\\
&= \int_0^1\int_0^1 xyH_1(x,1)h(x,1)\,dxdy && (4.35)\\
&\quad+\int_0^1\int_0^1 xH_1(1,y)h(1,y)\,dydx\\
&= \frac{1}{2}\int_0^1 xH_1(x,1)h(x,1)\,dx && (4.36)\\
&\quad+\frac{1}{2}\int_0^1 H_1(1,y)h(1,y)\,dy && (4.37)\\
&= \frac{1}{2}\int_0^1 xH_1(x,1)h(x,1)\,dx+\frac{1}{4}H_1^2(1,1)\\
&= -\frac{1}{2}\int_0^1 x^2H_1(x,1)H_{11}(x,1)\,dx+\frac{1}{4}H_1^2(1,1)\\
&= -\frac{1}{4}\int_0^1 x^2\,dH_1^2(x,1)+\frac{1}{4}H_1^2(1,1)\\
&= \frac{1}{2}\int_0^1 xH_1^2(x,1)\,dx.
\end{aligned}
$$

Following the same steps as above, we can show that

$$\int_0^1\int_0^1 yH_2(x,y)h(x,y)\,dxdy=\frac{1}{2}\int_0^1 yH_2^2(1,y)\,dy. \tag{4.38}$$

Now, Theorem 4.2.1 together with relations (4.32)–(4.38) yield

$$\theta^\tau=2\int_0^1 xH_1^2(x,1)\,dx+2\int_0^1 yH_2^2(1,y)\,dy-1. \tag{4.39}$$

Note that

$$H(x,1) = \lim_{u\to 0} \frac{C(ux,u)}{cu} \le \frac{x}{c} \;\text{ and }\; H(x,1) \ge H(x,x) = xH(1,1) = x. \qquad (4.40)$$

The first step is to find a decreasing density function f with support $(0,1)$ and an associated distribution function F in such a way as to minimize the objective function

$$J = \int_0^1 x f^2(x)\, dx$$

subject to the constraints that $c^{-1}x \ge F(x) \ge x$ for all $0 \le x \le 1$ (due to (4.40)) and that $\lim_{x\to 1} f(x) = \xi$, where $c \le 1$ and $\xi \in (0,1)$ are constants. We regard this as a problem of finding the minimal-cost trajectory from $x = 0, F = 0$ to $x = 1, F = 1$, which we approach by a Dynamic Programming argument.

Denote by $V(x,F)$ the following minimum

$$V(x,F) = \inf_{f\in\mathcal{F}} \left\{ \int_x^1 y f^2(y)\, dy \text{ subject to } \int_x^1 f(y)\, du = 1 - F \right\}.$$

Suppose we are starting from position (x_0, F_0). Further, consider a strategy which sets $f(x) = u$ for $x_0 \le x < x_0 + h$ and uses the optimal strategy for $x_0 + h \le x \le 1$. The cost of this strategy is

$$\int_{x_0}^{x_0+h} xu^2\, dx + V(x_0+h, F_0+uh) = x_0u^2h + V(x_0,F_0) + hV_1(x_0,F_0) + uhV_2(x_0,F_0) + o(h).$$

If we choose u optimally, we now have an optimal strategy from x_0 to 1; in other words,

$$V(x_0,F_0) = \inf_{u\in\mathcal{A}(x_0,F_0)} \left\{ V(x_0,F_0) + (x_0u^2 + V_1(x_0,F_0) + uV_2(x_0,F_0))h + o(h) \right\},$$

where V_1 and V_2 represent the partial derivatives of V and where $\mathcal{A}(x_0, F_0)$ represents the set of values u is permitted to take. This consists of $[0, f(x_0)]$ if (x_0, F_0) is in the interior of the accessible region, $[1, f(x_0)]$ if it is on the right-hand boundary, $[0, c^{-1}]$ if on the left-hand boundary.

As we let $h \to 0$, it can be seen that

$$\inf_{u \in \mathcal{A}(x_0, F_0)} \left\{ x_0 u^2 + V_1(x_0, F_0) + u V_2(x_0, F_0) \right\} = 0,$$

which is the optimality equation.

Minimizing over u, the optimal value u^* satisfies $u^*(x_0, F_0) = -\frac{1}{2x_0} V_2(x_0, F_0)$, as long as $u^* \in \mathcal{A}(x_0, F_0)$, in which case we conclude that $V_1(x_0, F_0) = \frac{1}{4x_0} V_2^2(x_0, F_0)$.

Let f be a feasible strategy and denote by V^f the associated value function $V^f(x_0, F_0) = \int_{x_0}^1 x f^2(x)\, dx$. If V^f satisfies the optimality equation and the associated boundary conditions, then f is the optimal strategy and $V = V^f$. Our approach, then, is to display the optimal strategy and to check that the optimality equation and boundary conditions are satisfied.

Define $k = -(1 - \xi)/\log(c\xi)$ and we show now that the optimal trajectory starting from $(0, 0)$ is

$$\begin{array}{llll} f(x) = 1/c & \text{and} & F(x) = x/c & \text{if } x < ck, \\ f(x) = k/x & \text{and} & F(x) = k + k\log x - k\log(ck) & \text{if } ck \le x \le k/\xi, \\ f(x) = \xi & \text{and} & F(x) = 1 - \xi(1 - x) & \text{if } k/\xi < x \le 1. \end{array} \tag{4.41}$$

Let D denote the triangular region bounded below by $F = x$ and above by $F = x/c$ and $F = 1 - \xi(1 - x)$. D therefore represents the set of points which are accessible from $(0, 0)$ and from which $(1, 1)$ is accessible without violating the restrictions. We divide D into sub-regions as follows:

- A is the region bounded below by $F = x$ and above by the curve $F = 1 + \xi \log x$.

- B is the region bounded above by $F = x/c$, below by $F = x$ and to the right by the curve $F = k - k\log(ck) + k\log x$, where $k = -(1 - \xi)/\log(c\xi)$.

- $C = D \cap (A \cup B)^c$.

In order to fully justify (4.41), the following claims will be shown:

(i) For $(x_0, F_0) \in A$, the trajectory which minimizes J, and the associated optimal value function, are $1 - F(x) = (1 - F_0)\frac{\log x}{\log x_0}$ and $V(x_0, F_0) = \frac{(1-F_0)^2}{-\log x_0}$, respectively;

(ii) For $(x_0, F_0) \in B$, the optimal strategy is to follow the trajectory $F(x) = F_0 + \frac{x_L}{c}\log\left(\frac{x}{x_0}\right)$ until it hits the point $(x_L, x_L/c)$, after which it follows the trajectory presented in (4.41). In addition, x_L is the solution of the equation

$$x_L = cF_0 + x_L \log(x_L/x_0), \tag{4.42}$$

and the optimal value function in region B is given by

$$V(x_0, F_0) = \frac{x_L^2}{c^2}\log\left(\frac{x_L}{x_0}\right) - \frac{x_L^2}{2c^2} + k(1-\xi) + \frac{1}{2}\xi^2.$$

(iii) For $(x_0, F_0) \in C$, the optimal strategy is to follow the trajectory $F(x) = F_0 + \xi x_U \log\left(\frac{x}{x_0}\right)$ until it hits the point $\big(x_U, 1 - \xi(1 - x_U)\big)$, after which it follows the trajectory presented in (4.41). In addition, x_U is the solution of the equation $\xi x_U + 1 - F_0 - \xi = \xi x_U \log(x_U/x_0)$, and the optimal value function in region C is given by $V(x_0, F_0) = \xi^2 x_U^2 \log\left(\frac{x_U}{x_0}\right) + \frac{1}{2}\xi^2(1 - x_U^2)$.

First of all, claim (i) does not claimed that the strategy is optimal. This is because the natural trajectory from (x_0, F_0) to $(1, 1)$, which is the one given in (4.41), arrives at $(1, 1)$ with $f(1-) > \xi$. In order to fit the criteria for acceptable trajectories, a small adjustment is required in the region of 1 so that $f(1-) = \xi$. The scale of the adjustment can be as small as desired, but it means that there is no optimal strategy, only a collection of ϵ-optimal strategies for any ϵ.

We first show claim (i). We begin by verifying that V and the proposed strategy satisfy the optimality equation. Note that

$$\frac{\partial V}{\partial F_0} = -2\frac{1-F_0}{-\log x_0}, \quad \frac{\partial V}{\partial x_0} = \frac{(1-F_0)^2}{(-\log x_0)^2} \cdot \frac{1}{x_0},$$

so that $V_2^2 = 4xV_1$, as required. One can check that $\frac{dF}{dx}\big|_{x=x_0} = -\frac{1}{2x_0}\frac{\partial V}{\partial F_0}$. f^* is non-increasing, since it takes the form constant$/x$.

Finally, we need to check that the optimal value of f is at least equal to 1 when (x_0, F_0) lies on the lower boundary of A, i.e., when $F_0 = x_0$. In this case $f^* = \frac{1-x_0}{-x_0 \log x_0} = y^{-1}(e^y - 1)$ if we write $x = e^{-y}$. Since we know that $e^y > 1 + y$, this is fine.

The proof of claim (ii) is less straightforward, as the quantity x_L, which features in the statement of the optimal strategy, is defined by an implicit equation (4.42). However, we have

$$\frac{1}{c}\frac{\partial x_L}{\partial F_0} = 1 + \frac{1}{c}\log(x_L/x_0)\frac{\partial x_L}{\partial F_0} + \frac{1}{c}\frac{\partial x_L}{\partial F_0}, \text{so that} \frac{\partial x_L}{\partial F_0} = -\frac{c}{\log(x_L/x_0)},$$

and

$$\frac{1}{c}\frac{\partial x_L}{\partial x_0} = \frac{1}{c}\log(x_L/x_0)\frac{\partial x_L}{\partial x_0} + \frac{1}{c}\frac{\partial x_L}{\partial x_0} - \frac{x_L}{cx_0}, \text{so that} \frac{\partial x_L}{\partial x_0} = \frac{x_L/x_0}{\log(x_L/x_0)},$$

Now, $\frac{\partial V}{\partial F_0} = 2\frac{x_L}{c^2}\log\left(\frac{x_L}{x_0}\right)\frac{\partial x_L}{\partial F_0} = -2\frac{x_L}{c}$ and $\frac{\partial V}{\partial x_0} = 2\frac{x_L}{c^2}\log\left(\frac{x_L}{x_0}\right)\frac{\partial x_L}{\partial x_0} - \frac{x_L^2}{c^2 x_0} = \frac{x_L^2}{c^2 x_0}$, and it is apparent that the optimality equation is satisfied. In addition, f is decreasing over this range and, at $x = x_0$, $\frac{dF}{dx}\big|_{x=x_0} = \frac{x_L}{cx_0} = -\frac{1}{2x_0}\frac{\partial V}{\partial F_0}$. On the lower boundary, where $x_0 = F_0$, we need to show that $f^* \geq 1$. But $f^* = x_L/(cx_0)$, and $c < 1$, $x_0 \leq x_L$, so that is fine. On the upper boundary, where $F_0 = x_0/c$, x_L is by definition equal to x_0, and $-\frac{1}{2x_0}V_2 = 1/c$, as required.

The proof of claim (iii) is very similar to the proof of claim (ii). We have

$$\xi\frac{\partial x_U}{\partial F_0} - 1 = \xi\log(x_U/x_0)\frac{\partial x_U}{\partial F_0} + \xi\frac{\partial x_U}{\partial F_0}, \text{so that} \frac{\partial x_U}{\partial F_0} = -\frac{1}{\xi\log(x_U/x_0)},$$

and

$$\xi\frac{\partial x_U}{\partial x_0} = \xi\log(x_U/x_0)\frac{\partial x_U}{\partial x_0} + \xi\frac{\partial x_U}{\partial x_0} - \frac{\xi x_U}{x_0}, \text{so that} \frac{\partial x_U}{\partial x_0} = \frac{x_U/x_0}{\log(x_U/x_0)},$$

Now, $\frac{\partial V}{\partial F_0} = 2\xi^2 x_U \log\left(\frac{x_L}{x_0}\right)\frac{\partial x_U}{\partial F_0} = -2\xi x_U$ and $\frac{\partial V}{\partial x_0} = 2\xi^2 x_U \log\left(\frac{x_U}{x_0}\right)\frac{\partial x_U}{\partial x_0} - \xi^2\frac{x_U^2}{x_0} = \xi^2\frac{x_U^2}{x_0}$, and it is apparent that the optimality equation is satisfied. The checks on the boundaries proceed as before.

We have demonstrated the optimal strategy throughout the region D, and can therefore state that

$$\begin{aligned} V(0,0) &= \int_0^{ck} c^{-2}x\,dx + \int_{ck}^{k/\xi}\frac{k^2}{x}\,dx + \int_{k/\xi}^{1}\xi^2 x\,dx \\ &= \frac{k^2}{2} - k^2\log(c\xi) + \frac{1}{2}(\xi^2 - k^2) \\ &= \frac{\xi^2}{2} - \frac{(1-\xi)^2}{\log(c\xi)}. \end{aligned}$$

This quantity represents the minimal value of $\int_0^1 xH_1^2(x,1)\,dx$ under the restrictions that $x \le H(x,1) \le x/c$ and $H_1(1-,1) = \xi$. For $\int_0^1 xH^2(1,x)\,dx$ we perform the same minimization, with the exception that ξ is replaced by $1-\xi$. This shows us that

$$\theta^\tau \ge -1 + 2\inf_{\xi\in(0,1)}\left\{\frac{\xi^2}{2} - \frac{(1-\xi)^2}{\log(c\xi)} + \frac{(1-\xi)^2}{2} - \frac{\xi^2}{\log(c(1-\xi))}\right\}.$$

The minimum occurs at $\xi = \frac{1}{2}$, giving a minimal value of

$$-1 + 2\left(\frac{1}{4} - \frac{1}{2\log(c/2)}\right) = -\frac{1}{2} - \frac{1}{\log(c/2)}.$$

$\square$

Proof of Theorem 4.2.3. Clearly,

$$\begin{aligned}
& \int_0^1 \int_0^1 H(x,y)h(x,y)\,dxdy \\
= \;& \tfrac{1}{4}\sum_{i=1}^m \sum_{j=1}^m c_i c_j \alpha_j \beta_j \tfrac{1}{(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \sum_{i=1}^m \tfrac{c_i^2}{4} + \sum_{i\neq j} c_i c_j \alpha_j \beta_j \tfrac{1}{(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \tfrac{(\sum_{i=1}^m c_i)^2}{4} - \sum_{i\neq j} \tfrac{c_i c_j}{4} + \sum_{i\neq j} c_i c_j \alpha_j \beta_j \tfrac{1}{(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \tfrac{1}{4} + \sum_{i\neq j} c_i c_j \tfrac{4\alpha_j\beta_j - (\alpha_i+\alpha_j)(\beta_i+\beta_j)}{4(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \tfrac{1}{4} + \sum_{i\neq j} c_i c_j \tfrac{2\alpha_i\beta_i + 2\alpha_j\beta_j - (\alpha_i+\alpha_j)(\beta_i+\beta_j)}{4(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \tfrac{1}{4} + \sum_{i\neq j} c_i c_j \tfrac{(\alpha_i-\alpha_j)(\beta_i-\beta_j)}{4(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
= \;& \tfrac{1}{4} + \sum_{i\neq j} c_i c_j \tfrac{-(\alpha_i-\alpha_j)^2}{4(\alpha_i+\alpha_j)(\beta_i+\beta_j)} \\
\leq \;& \tfrac{1}{4}.
\end{aligned}$$

Thus, the latter and Theorem 4.2.1 illustrate that $\theta^\tau \leq 0$. □

Proof of Theorem 4.2.4. Put

$$\begin{cases}
\theta_n = E\left\{ sgn\big((U_1-U_2)(V_1-V_2)\big) I\big(\max(U_1,V_1,U_2,V_2) \leq \tfrac{k}{n}\big)\right\}, \\
\tilde{h}(u_1,v_1,u_2,v_2) = sgn\big((u_1-u_2)(v_1-v_2)\big) I\big(\max(u_1,v_1,u_2,v_2) \leq \tfrac{k}{n}\big) - \theta_n, \\
\tilde{h}_1(u_1,v_1) = E\left\{ sgn\big((u_1-U_2)(v_1-V_2)\big) I\big(\max(u_1,v_1,U_2,V_2) \leq \tfrac{k}{n}\big)\right\} - \theta_n, \\
S_{1n} = \sum_{i=1}^n \tilde{h}_1(U_i,V_i), \\
S_{2n} = \sum_{1\leq i<j\leq n} \left\{ \tilde{h}(U_i,V_i,U_j,V_j) - \tilde{h}_1(U_i,V_i) - \tilde{h}_1(U_j,V_j)\right\}, \\
Z_n = \tfrac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \left\{ sgn\big((U_i-U_j)(V_i-V_j)\big) I\big(\max(U_i,V_i,U_j,V_j) \leq \tfrac{k}{n}\big) - \theta_n \right\}.
\end{cases}$$

Then it follows from the Hoeffding decomposition (Hoeffding (1948) or Lemma A from page 178 of Serfling (1980)) that

$$Z_n = \frac{2}{n} S_{1n} + \frac{2}{n(n-1)} S_{2n}. \tag{4.43}$$

In addition, Lemma A from page 183 of Serfling (1980) leads to

$$EZ_n^2 = \frac{4(n-2)}{n(n-1)} E\tilde{h}_1^2(U_1, V_1) + \frac{2}{n(n-1)} E\tilde{h}^2(U_1, V_1, U_2, V_2). \tag{4.44}$$

It is straightforward to check that

$$\theta_n / C^2\left(\frac{k}{n}, \frac{k}{n}\right) \to \theta^\tau \tag{4.45}$$

and

$$\begin{aligned}
\tilde{h}_1(u_1, v_1) &= 2P\big(u_1 > U_2, v_1 > V_2, \max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) \\
&\quad +2P\big(u_1 < U_2, v_1 < V_2, \max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) \\
&\quad -P\big(\max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) - \theta_n \\
&= 4P\big(u_1 > U_2, v_1 > V_2, \max(u_1, v_1, U_2, V_2\big) \leq \tfrac{k}{n}) \\
&\quad -2P\big(u_1 > U_2, \max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) \\
&\quad -2P\big(v_1 > V_2, \max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) \\
&\quad +P\big(\max(u_1, v_1, U_2, V_2) \leq \tfrac{k}{n}\big) - \theta_n \\
&= 4C(u_1, v_1) I\big(\max(u_1, v_1) \leq \tfrac{k}{n}\big) \\
&\quad -2C\big(u_1, \tfrac{k}{n}\big) I\big(\max(u_1, v_1) \leq \tfrac{k}{n}\big) \\
&\quad -2C\big(\tfrac{k}{n}, v_1\big) I\big(\max(u_1, v_1) \leq \tfrac{k}{n}\big) \\
&\quad +C\big(\tfrac{k}{n}, \tfrac{k}{n}\big) I\big(\max(u_1, v_1) \leq \tfrac{k}{n}\big) - \theta_n.
\end{aligned}$$

Thus, it follows from Assumption 4.2.1 that

$$\begin{aligned}
&\frac{E\tilde{h}_1^2(U_1,V_1)}{C^3\left(\frac{k}{n},\frac{k}{n}\right)} \\
= \quad & C^{-3}\left(\tfrac{k}{n},\tfrac{k}{n}\right) E\Big\{ 16C^2(U_1,V_1) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& +4C^2\left(U_1,\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) + 4C^2\left(\tfrac{k}{n},V_1\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& +C^2\left(\tfrac{k}{n},\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) + \theta_n^2 \\
& -16C(U_1,V_1)C\left(U_1,\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& -16C(U_1,V_1)C\left(\tfrac{k}{n},V_1\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& +8C(U_1,V_1)C\left(\tfrac{k}{n},\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& +8C\left(U_1,\tfrac{k}{n}\right) C\left(\tfrac{k}{n},V_1\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& -4C\left(U_1,\tfrac{k}{n}\right) C\left(\tfrac{k}{n},\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& -4C\left(\tfrac{k}{n},V_1\right) C\left(\tfrac{k}{n},\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& -2\theta_n\Big(4C(U_1,V_1) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) - 2C\left(U_1,\tfrac{k}{n}\right) I\left(\max(U_1,V_1)\le \tfrac{k}{n}\right) \\
& -2C\left(\tfrac{k}{n},U_1\right) I\left(\max(V_1,U_1)\le \tfrac{k}{n}\right) + C\left(\tfrac{k}{n},\tfrac{k}{n}\right) I\left(\max(V_1,U_1)\le \tfrac{k}{n}\right)\Big)\Big\} \\
\to \quad & 16\int_0^1\int_0^1 H^2(x,y)\,dH(x,y) - 16\int_0^1\int_0^1 H(x,1)H(x,y)\,dH(x,y) \\
& -16\int_0^1\int_0^1 H(1,y)H(x,y)\,dH(x,y) + 8\int_0^1\int_0^1 H(x,y)\,dH(x,y) \\
& +8\int_0^1\int_0^1 H(x,1)H(1,y)\,dH(x,y) - \tfrac{1}{3}
\end{aligned} \tag{4.46}$$

and

$$\frac{E\tilde{h}^2(U_1,V_1,U_2,V_2)}{C^2\left(\frac{k}{n},\frac{k}{n}\right)} = \frac{\theta_n}{C^2\left(\frac{k}{n},\frac{k}{n}\right)} + o(1) \to 1. \tag{4.47}$$

By equations (4.44), (4.46) and (4.47), and the fact that $nC\left(\frac{k}{n},\frac{k}{n}\right)\to\infty$, we have

$$E\left(Z_n/C^2\left(\frac{k}{n},\frac{k}{n}\right)\right)^2 \to 0,$$

which in turn implies that $Z_n/C^2\left(\frac{k}{n},\frac{k}{n}\right) \xrightarrow{p} 0$. Hence, (4.45) allows us to conclude that

$$\frac{2}{n(n-1)C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sum_{1\le i<j\le n} sgn\big((U_i-U_j)(V_i-V_j)\big)I\left(\max(U_i,V_i,U_j,V_j)\le\frac{k}{n}\right)\xrightarrow{p}\theta^\tau. \tag{4.48}$$

Denote $G_{n1}(x)=\frac{1}{n+1}\sum_{i=1}^n I(U_i\le x)$ and $G_{n2}(y)=\frac{1}{n+1}\sum_{i=1}^n I(V_i\le y)$. Note that

$$\begin{aligned} & sgn\big((\hat U_i-\hat U_j)(\hat V_i-\hat V_j)\big)I\big(\max(\hat U_i,\hat V_i,\hat U_j,\hat U_j)\le\tfrac{k}{n}\big)\\ = \;& sgn\big((U_i-U_j)(V_i-V_j)\big)I\left(\max(U_i,U_j)\le G_{n1}^-(\tfrac{k}{n}),\max(V_i,V_j)\le G_{n2}^-(\tfrac{k}{n})\right),\end{aligned}$$

$\frac{n}{k}G_{n1}^-\big(\frac{k}{n}\big)\xrightarrow{p}1$ and $\frac{n}{k}G_{n2}^-\big(\frac{k}{n}\big)\xrightarrow{p}1$. These properties, equation (4.48) and the continuity of H yield

$$\frac{2}{n(n-1)C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sum_{1\le i<j\le n} sgn\big((\hat U_i-\hat U_j)(\hat V_i-\hat V_j)\big)I\left(\max(\hat U_i,\hat V_i,\hat U_j,\hat V_j)\le\frac{k}{n}\right)\xrightarrow{p}\theta^\tau. \tag{4.49}$$

Similarly, we can show that

$$\frac{2}{n(n-1)C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sum_{1\le i<j\le n} I\left(\max(\hat U_i,\hat V_i,\hat U_j,\hat V_j)\le\frac{k}{n}\right)\xrightarrow{p}1. \tag{4.50}$$

Therefore, it follows from (4.49) and (4.50) that $\hat\theta^\tau(k)\xrightarrow{p}\theta^\tau$. □

Proof of Theorem 4.2.5. It is worth mentioning that the current proof follows the same notations defined in the proof of Theorem 4.2.4. In addition, we define

$$\begin{aligned}\beta_{n1}(x,y) \;&= \frac{2}{n(n-1)C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sum\nolimits_{1\le i<j\le n}\\ & sgn\big((U_i-U_j)(V_i-V_j)\big)I\left(\max(U_i,U_j)\le\tfrac{k}{n}x\right)I\left(\max(V_i,V_j)\le\tfrac{k}{n}y\right)\end{aligned}$$

and

$$\beta_{n2}(x,y)=\frac{2}{n(n-1)C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sum_{1\le i<j\le n} I\left(\max(U_i,U_j)\le\frac{k}{n}x\right)I\left(\max(V_i,V_j)\le\frac{k}{n}y\right).$$

Now, Assumption 4.2.4 leads to

$$\begin{array}{rl} & A^{-1}\left(\frac{k}{n}\right)\left\{E\beta_{n1}(x,y) - 4\int_0^x\int_0^y H(s,t)H_{12}(s,t)\,dtds + H^2(x,y)\right\} \\ \to & 4\int_0^x\int_0^y Q(s,t)H_{12}(s,t)\,dtds + 4\int_0^x\int_0^y H(s,t)q(s,t)\,dtds - 2H(x,y)Q(x,y) \end{array} \tag{4.51}$$

and

$$A^{-1}\left(\frac{k}{n}\right)\left\{E\beta_{n2}(x,y) - H^2(x,y)\right\} \to 2H(x,y)Q(x,y). \tag{4.52}$$

By (4.43), (4.44), (4.46), (4.47) and the fact that $nC\left(\frac{k}{n},\frac{k}{n}\right) \to \infty$, we have

$$\sqrt{nC\left(\frac{k}{n},\frac{k}{n}\right)}\left\{\beta_{n1}(1,1) - E\beta_{n1}(1,1)\right\} = \frac{2\sigma_1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}} + o_p(1), \tag{4.53}$$

where σ_1^2 is defined in (4.14). Similarly,

$$\sqrt{nC\left(\frac{k}{n},\frac{k}{n}\right)}\{\beta_{n2}(1,1) - E\beta_{n2}(1,1)\} = \frac{2}{\sqrt{n}}\sum_{i=1}^n \frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\hat{h}_1^2(U_1,V_1)}} + o_p(1), \tag{4.54}$$

where $\hat{h}_1(u_1,v_1) = I\big(\max(u_1,v_1) \le \frac{k}{n}\big) - C\big(\frac{k}{n},\frac{k}{n}\big)$. Using $H(1,1) = 1$ and $Q(1,1) = 0$, and (4.51)–(4.54), we have

$$\begin{array}{rl} & \sqrt{nC\left(\frac{k}{n},\frac{k}{n}\right)}\left\{\frac{\beta_{n1}(1,1)}{\beta_{n2}(1,1)} - \theta^\tau\right\} \\ = & \sqrt{nC\left(\frac{k}{n},\frac{k}{n}\right)}\left\{\frac{\beta_{n1}(1,1)-E\beta_{n1}(1,1)}{\beta_{n2}(1,1)} - \frac{(\beta_{n2}(1,1)-E\beta_{n2}(1,1))E\beta_{n1}(1,1)}{\beta_{n2}(1,1)E\beta_{n2}(1,1)}\right\} \\ & +\sqrt{nC\left(\frac{k}{n},\frac{k}{n}\right)}\left\{\frac{E\beta_{n1}(1,1)}{E\beta_{n2}(1,1)} - \theta^\tau\right\} \\ = & \frac{2\sigma_1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}} - \frac{2\theta^\tau}{\sqrt{n}}\sum_{i=1}^n \frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\hat{h}_1^2(U_1,V_1)}} \\ & +\lambda\left\{4\int_0^1\int_0^1 Q(s,t)H_{12}(s,t)\,dtds + 4\int_0^1\int_0^1 H(s,t)q(s,t)\,dtds\right\} + o_p(1). \end{array} \tag{4.55}$$

Further, we have

$$
\begin{aligned}
&\sqrt{nC\big(\tfrac{k}{n},\tfrac{k}{n}\big)}\big\{\hat{\theta}(k)-\theta^{\tau}\big\} \\
= \;& \sqrt{nC\big(\tfrac{k}{n},\tfrac{k}{n}\big)}\bigg\{\frac{\beta_{n1}\Big(\frac{n}{k}G_{n1}^{-}\big(\frac{k}{n}\big),\frac{n}{k}G_{n2}^{-}\big(\frac{k}{n}\big)\Big)}{\beta_{n2}\Big(\frac{n}{k}G_{n1}^{-}\big(\frac{k}{n}\big),\frac{n}{k}G_{n2}^{-}\big(\frac{k}{n}\big)\Big)}-4\int_0^{\frac{n}{k}G_{n1}^{-}\left(\frac{k}{n}\right)}\int_0^{\frac{n}{k}G_{n2}^{-}\left(\frac{k}{n}\right)} H(s,t)H_{12}(s,t)\,dtds \\
& +H^2\Big(\tfrac{n}{k}G_{n1}^{-}\big(\tfrac{k}{n}\big),\tfrac{n}{k}G_{n2}^{-}\big(\tfrac{k}{n}\big)\Big)\bigg\} \\
& +4\sqrt{nC\big(\tfrac{k}{n},\tfrac{k}{n}\big)}\bigg\{\int_0^{\frac{n}{k}G_{n1}^{-}\left(\frac{k}{n}\right)}\int_0^{\frac{n}{k}G_{n2}^{-}\left(\frac{k}{n}\right)} H(s,t)H_{12}(s,t)\,dtds \\
& -\int_0^1\int_0^1 H(s,t)H_{12}(s,t)\,dtds\bigg\} \\
& -\sqrt{nC\big(\tfrac{k}{n},\tfrac{k}{n}\big)}\Big\{H^2\Big(\tfrac{n}{k}G_{n1}^{-}\big(\tfrac{k}{n}\big),\tfrac{n}{k}G_{n2}^{-}\big(\tfrac{k}{n}\big)\Big)-H^2(1,1)\Big\} \\
= \;& \sqrt{nC\big(\tfrac{k}{n},\tfrac{k}{n}\big)}\,\Big\{\tfrac{\beta_{n1}(1,1)}{\beta_{n2}(1,1)}-\theta^{\tau}\Big\} \\
& +4\sqrt{c}\sqrt{k}\big\{\tfrac{n}{k}G_{n1}^{-}\big(\tfrac{k}{n}\big)-1\big\}\int_0^1 H(1,t)H_{12}(1,t)\,dt \\
& +4\sqrt{c}\sqrt{k}\big\{\tfrac{n}{k}G_{n2}^{-}\big(\tfrac{k}{n}\big)-1\big\}\int_0^1 H(s,1)H_{12}(s,1)\,ds \\
& -2\sqrt{c}\sqrt{k}\big\{\tfrac{n}{k}G_{n1}^{-}\big(\tfrac{k}{n}\big)-1\big\}H_1(1,1) \\
& -2\sqrt{c}\sqrt{k}\big\{\tfrac{n}{k}G_{n2}^{-}\big(\tfrac{k}{n}\big)-1\big\}H_2(1,1)+o_p(1).
\end{aligned}
\tag{4.56}
$$

It is not difficult to find that

$$
\left\{
\begin{array}{l}
E\left\{\frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}\frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\hat{h}_1^2(U_1,V_1)}}\right\}=\frac{\theta_n-\theta_n C\left(\frac{k}{n},\frac{k}{n}\right)}{\sigma_1 C^2\left(\frac{k}{n},\frac{k}{n}\right)\left(1+o(1)\right)}\to\frac{\theta^{\tau}}{\sigma_1} \\
E\left\{\frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}\frac{I\left(U_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\right\}=\frac{\theta_n-\frac{k}{n}\theta_n}{\sigma_1 C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sqrt{\frac{n}{k}C\big(\frac{k}{n},\frac{k}{n}\big)}\{1+o(1)\}\to\frac{\theta^{\tau}\sqrt{c}}{\sigma_1} \\
E\left\{\frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}\frac{I\left(V_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\right\}=\frac{\theta_n-\frac{k}{n}\theta_n}{\sigma_1 C^2\left(\frac{k}{n},\frac{k}{n}\right)}\sqrt{\frac{n}{k}C\big(\frac{k}{n},\frac{k}{n}\big)}\{1+o(1)\}\to\frac{\theta^{\tau}\sqrt{c}}{\sigma_1} \\
E\left\{\frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}\frac{I\left(U_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\right\}=\frac{C\left(\frac{k}{n},\frac{k}{n}\right)-\frac{k}{n}C\left(\frac{k}{n},\frac{k}{n}\right)}{C\left(\frac{k}{n},\frac{k}{n}\right)}\sqrt{\frac{n}{k}C\big(\frac{k}{n},\frac{k}{n}\big)}\{1+o(1)\}\to\sqrt{c} \\
E\left\{\frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}\frac{I\left(V_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\right\}=\frac{C\left(\frac{k}{n},\frac{k}{n}\right)-\frac{k}{n}C\left(\frac{k}{n},\frac{k}{n}\right)}{C\left(\frac{k}{n},\frac{k}{n}\right)}\sqrt{\frac{n}{k}C\big(\frac{k}{n},\frac{k}{n}\big)}\{1+o(1)\}\to\sqrt{c} \\
E\left\{\frac{I\left(U_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\frac{I\left(V_i\le\frac{k}{n}\right)-\frac{k}{n}}{\sqrt{k/n}}\right\}=\frac{C\left(\frac{k}{n},\frac{k}{n}\right)-\left(\frac{k}{n}\right)^2}{k/n}\to c.
\end{array}
\right.
$$

Consequently, using the Cramér-device, we can show that

$$
\begin{aligned}
& \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\hat{h}_1^2(U_1,V_1)}}, \sqrt{k}(\tfrac{n}{k}G_{n1}^{-}(\tfrac{k}{n})-1), \sqrt{k}(\tfrac{n}{k}G_{n2}^{-}(\tfrac{k}{n})-1)\right)^T \\
= & \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\tilde{h}_1(U_i,V_i)}{\sqrt{E\tilde{h}_1^2(U_1,V_1)}}, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\hat{h}_1(U_i,V_i)}{\sqrt{E\hat{h}_1^2(U_1,V_1)}},\right. \\
& \left.-\frac{1}{\sqrt{k}}\sum_{i=1}^{n}(I(U_i\le\tfrac{k}{n})-\tfrac{k}{n}), -\frac{1}{\sqrt{k}}\sum_{i=1}^{n}(I(V_i\le\tfrac{k}{n})-\tfrac{k}{n})\right)^T + o_p(1) \\
\xrightarrow{d} & \; N(0,\Sigma)
\end{aligned} \tag{4.57}
$$

as $n \to \infty$, where

$$
\Sigma = \begin{pmatrix} 1 & \frac{\theta^\tau}{\sigma_1} & -\frac{\theta^\tau\sqrt{c}}{\sigma_1} & -\frac{\theta^\tau\sqrt{c}}{\sigma_1} \\ -\frac{\theta^\tau}{\sigma_1} & 1 & -\sqrt{c} & -\sqrt{c} \\ -\frac{\theta^\tau\sqrt{c}}{\sigma_1} & -\sqrt{c} & 1 & c \\ -\frac{\theta^\tau\sqrt{c}}{\sigma_1} & -\sqrt{c} & c & 1 \end{pmatrix}.
$$

Therefore, it follows from equations (4.55)–(4.57) that (4.12) holds. □

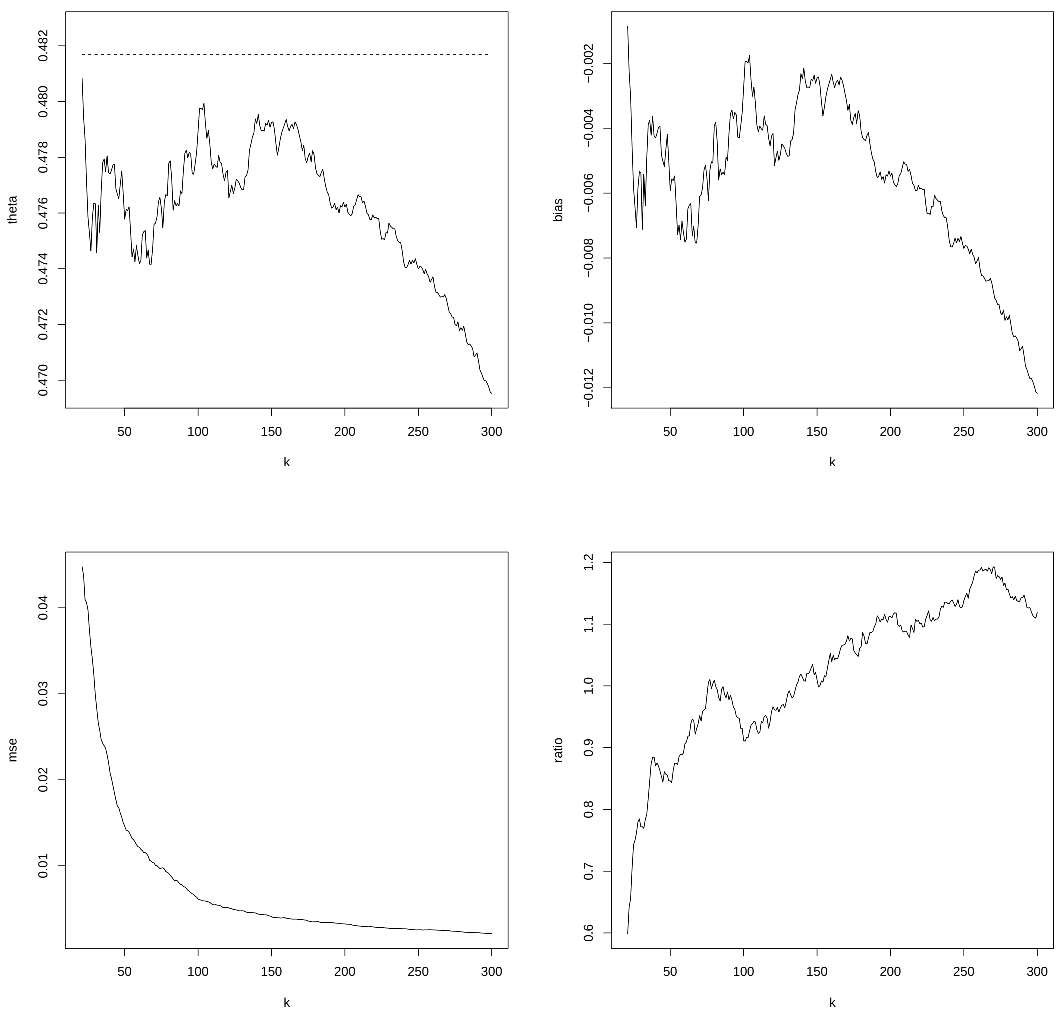


Figure 4.4: The estimator $\hat{\theta}^{\tau}(k)$, its bias, mean squared error and ratio of asymptotic variance to the bootstrap estimator are plotted against $k = 21, \cdots, 300$ for t copula with $\rho = 0.5$ and $\nu = 1$ given in Example 4.3.2 of Section 4.3.

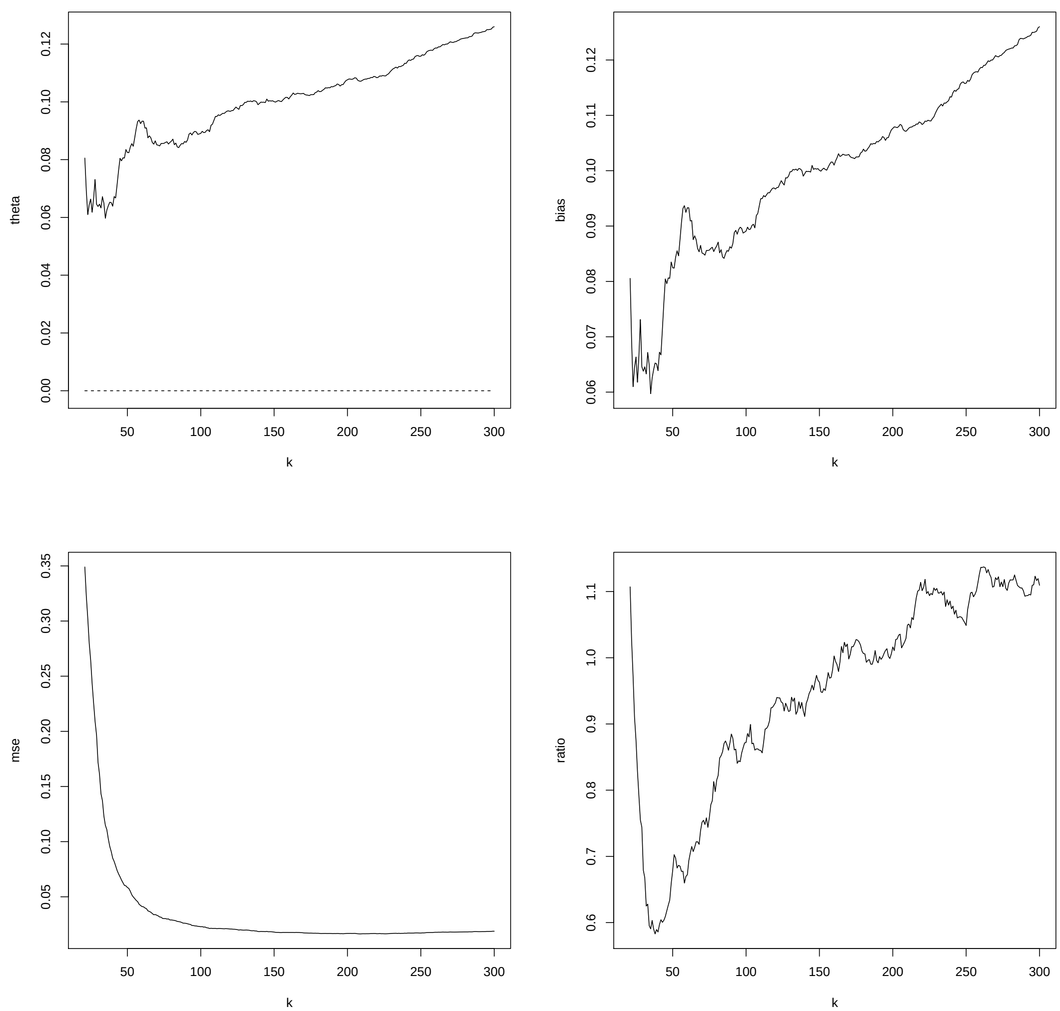


Figure 4.5: The estimator $\hat{\theta}^{\tau}(k)$, its bias, mean squared error and ratio of asymptotic variance to the bootstrap estimator are plotted against $k = 21, \cdots, 300$ for normal copula with $\rho = 0.5$ given in Example 4.3.5 of Section 4.3.

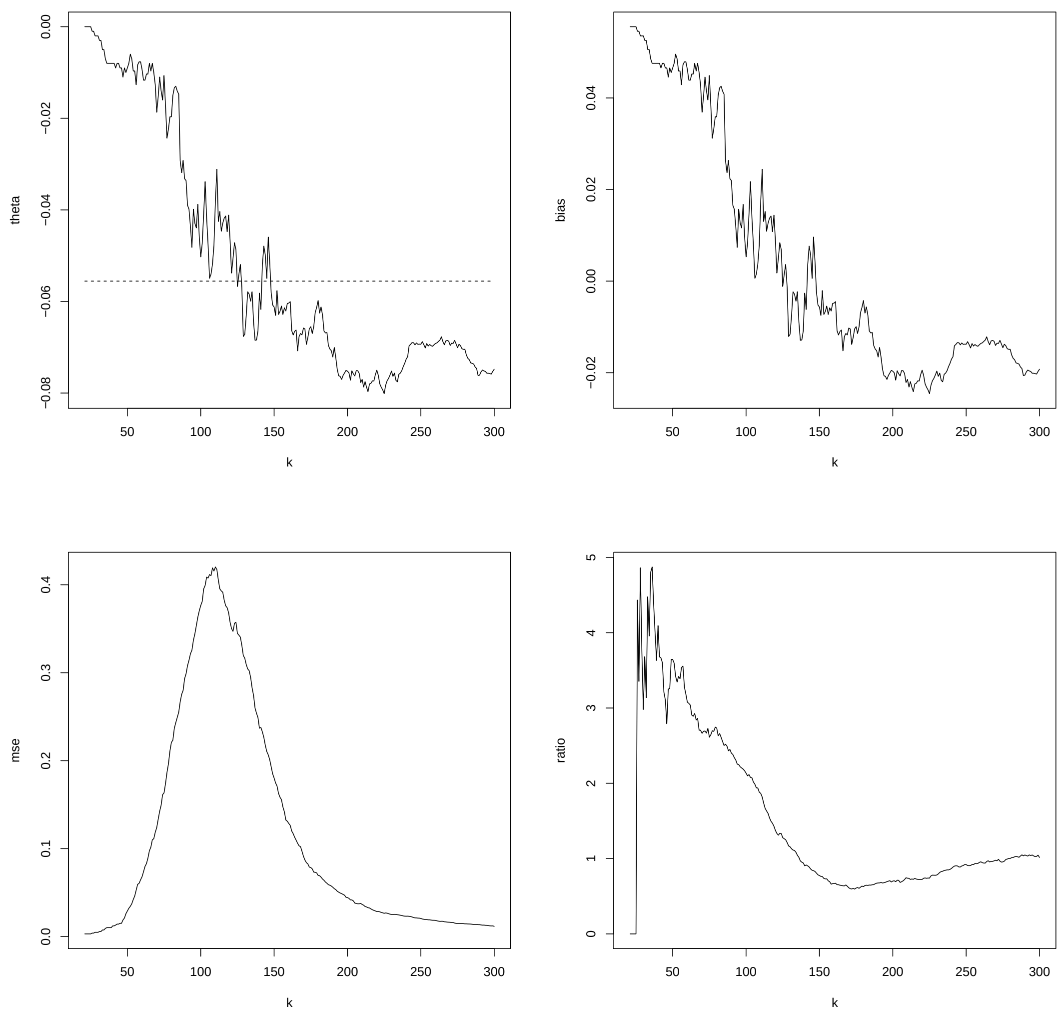


Figure 4.6: The estimator $\hat{\theta}^{\tau}(k)$, its bias, mean squared error and ratio of asymptotic variance to the bootstrap estimator are plotted against $k = 21, \cdots, 300$ for Farlie-Gumbel-Morgenstern copula with $\xi = -1$ and given in Example 4.3.6 of Section 4.3.

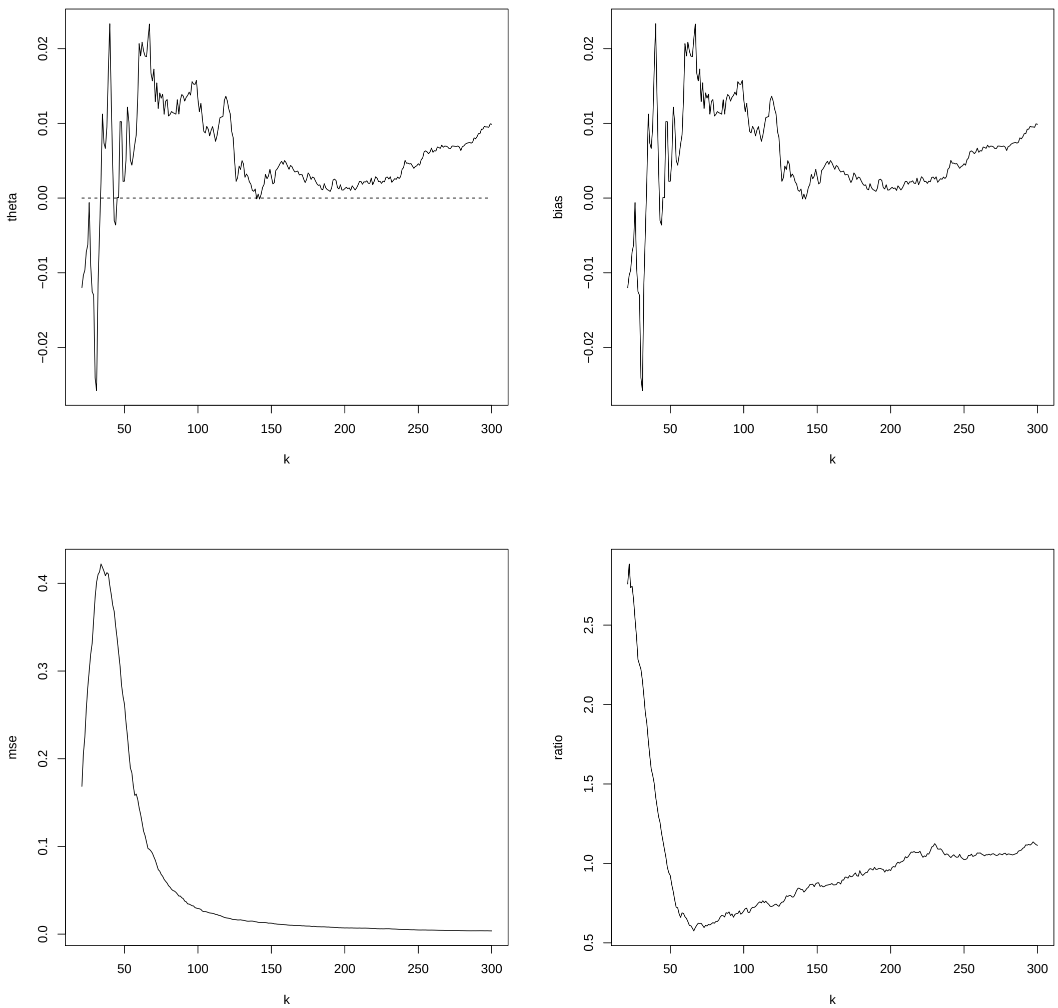


Figure 4.7: The estimator $\hat{\theta}^{\tau}(k)$, its bias, mean squared error and ratio of asymptotic variance to the bootstrap estimator are plotted against $k = 21, \cdots, 300$ for Farlie-Gumbel-Morgenstern copula with $\xi = 1$ and given in Example 4.3.6 of Section 4.3.

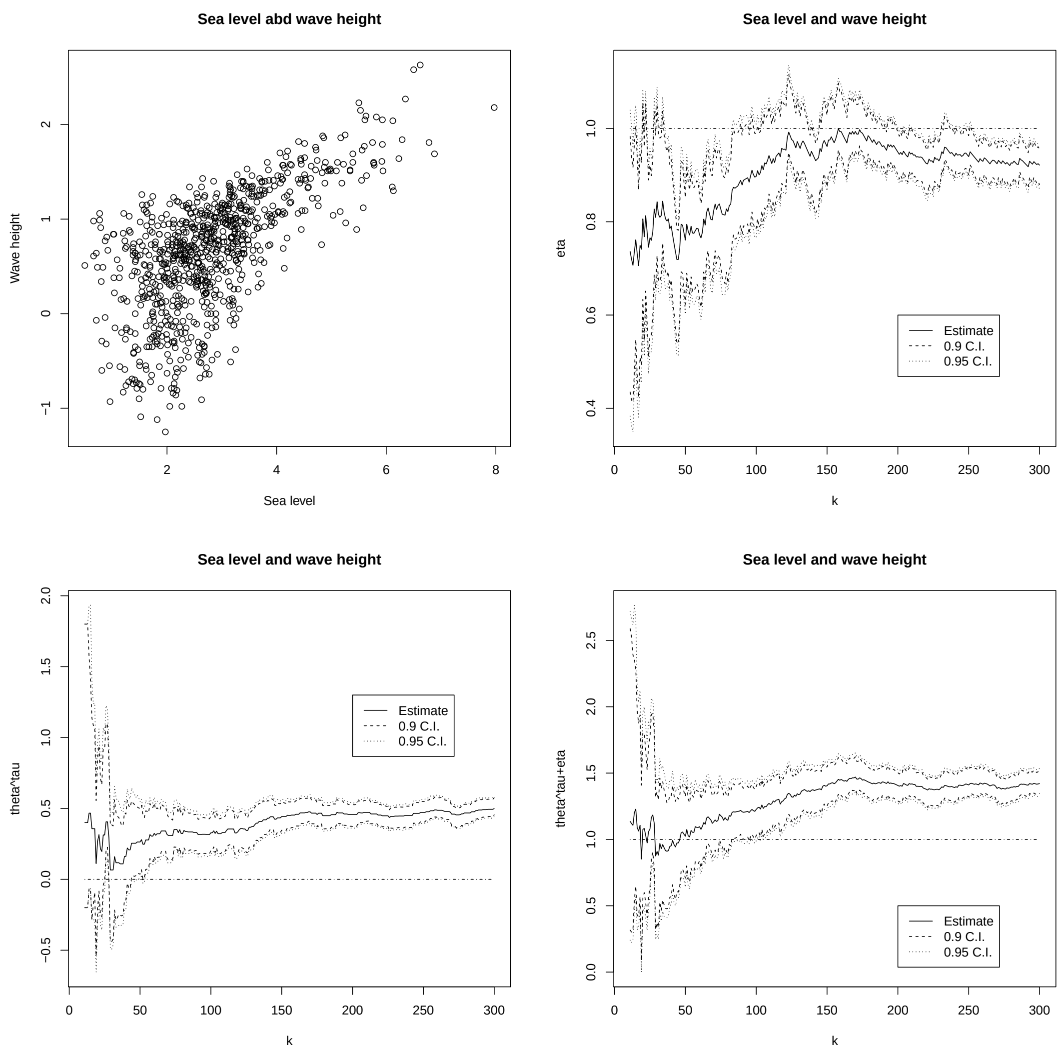


Figure 4.8: Sea level and wave height. Estimators $\hat{\eta}(k)$, $\hat{\theta}^{\tau}(k)$, $\hat{\theta}^{\tau}(k) + \hat{\eta}(k)$, and their intervals with level 0.9 and 0.95 are plotted against k.

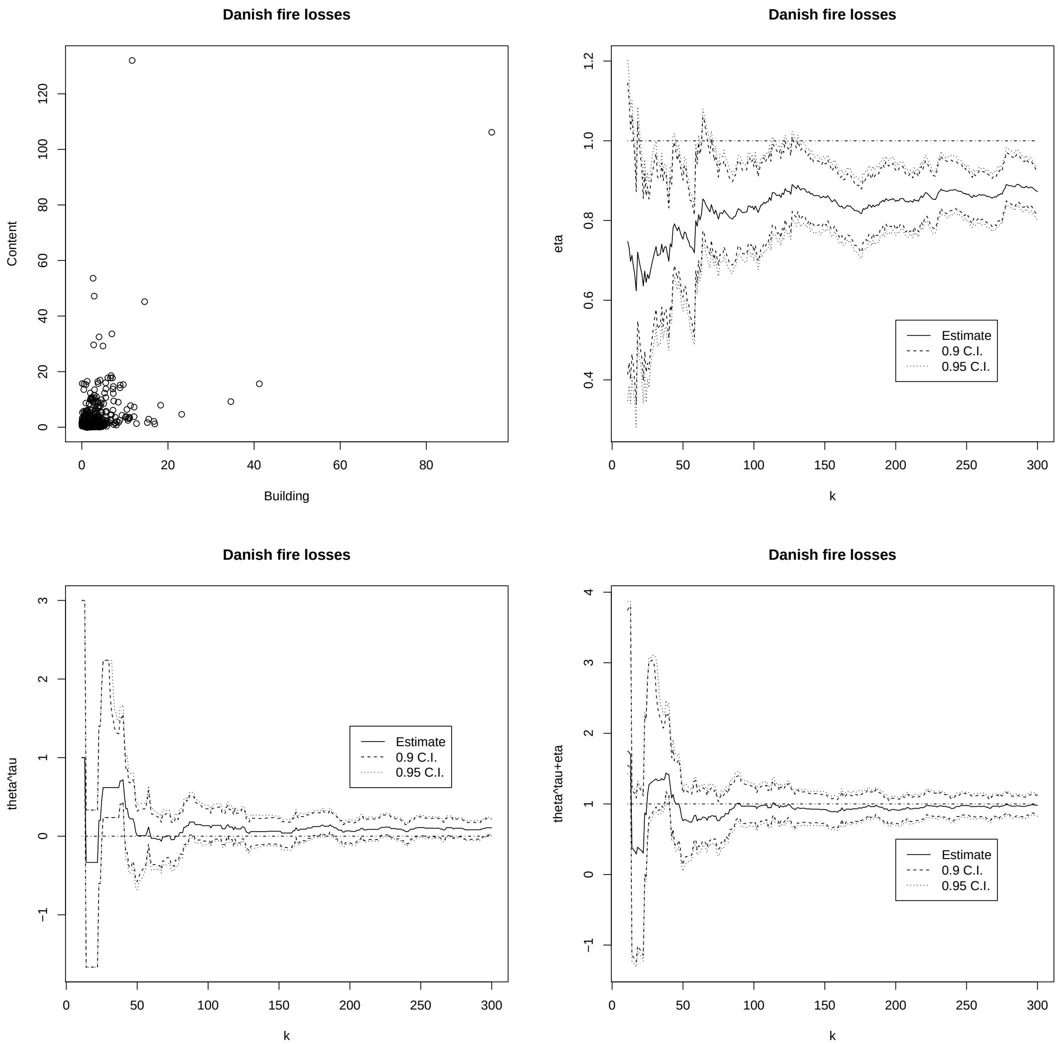


Figure 4.9: Danish fire losses. Estimators $\hat{\eta}(k)$, $\hat{\theta}^{\tau}(k)$, $\hat{\theta}^{\tau}(k) + \hat{\eta}(k)$, and their intervals with level 0.9 and 0.95 are plotted against k.

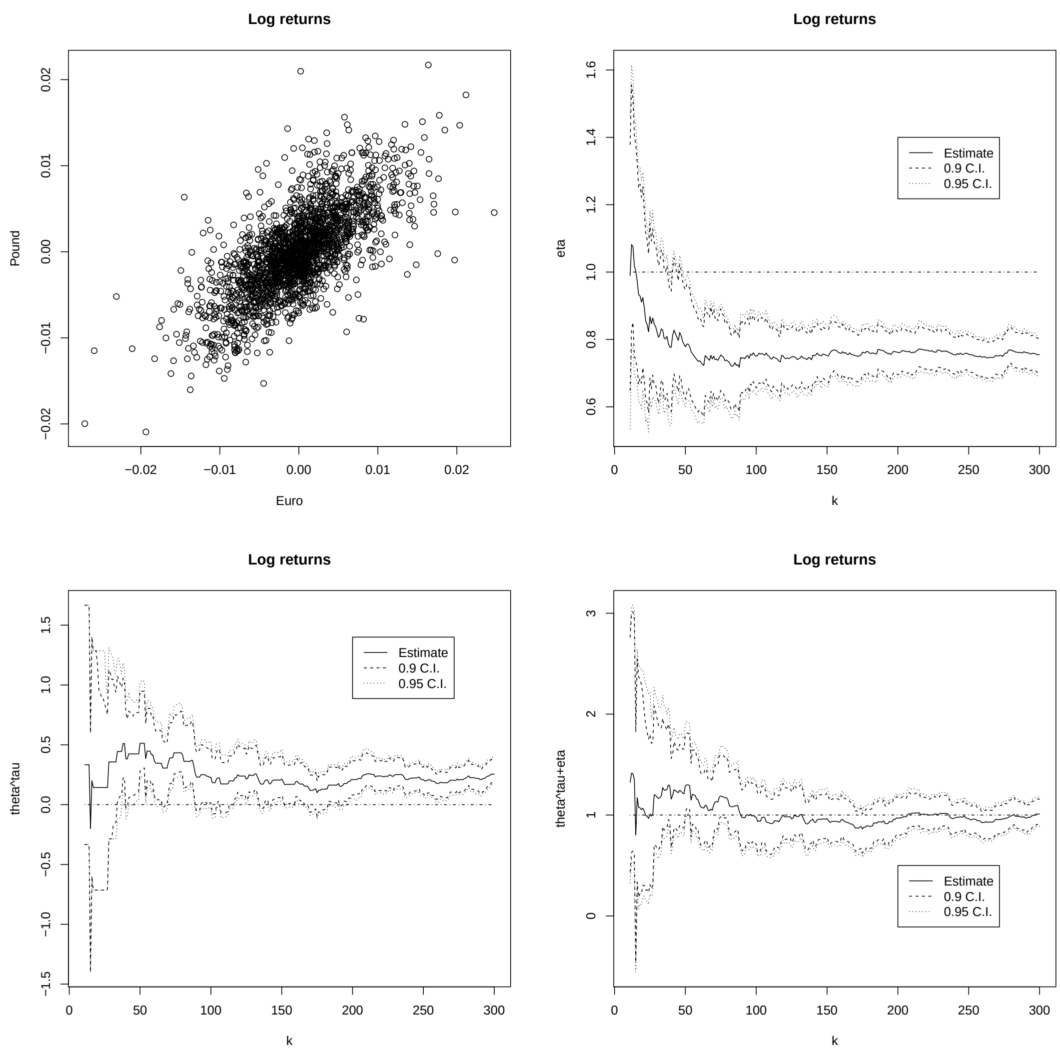


Figure 4.10: Log returns of exchange rates. Estimators $\hat{\eta}(k)$, $\hat{\theta}^{\tau}(k)$, $\hat{\theta}^{\tau}(k)+\hat{\eta}(k)$, and their intervals with level 0.9 and 0.95 are plotted against k.

CHAPTER 5

INTERVAL ESTIMATION FOR THE NEW TAIL DEPENDENCE MEASURE

Systemic risk concerns extreme co-movement of several financial variables, which involves characterizing tail dependence. The coefficient of tail dependence was proposed by [53, 54] to distinguish asymptotic independence and asymptotic dependence. Recently a new measure based on the conditional Kendall's tau was proposed by [104] to measure the tail dependence and to distinguish asymptotic independence and asymptotic dependence. For effectively constructing a confidence interval for this new measure, in this chapter we propose a smooth jackknife empirical likelihood method, which does not need to estimate any additional quantities such as asymptotic variance. A simulation study shows that the proposed method has a good finite sample performance. This chapter is based on A. Liu, Y. Hou and L. Peng (2015). Interval Estimation for a Measure of Tail Dependence. *Insurance: Mathematics and Economics 64, 294-305.*

5.1 Introduction

A recent research interest in risk management focuses on systemic risk in banking industry and insurance companies. Systemic risk concerns extreme co-movements of key financial variables. Effectively measuring tail dependence plays an important role in understanding and managing systemic risk. See [105] for measuring systemic risk and using it to predict future economic downturns; [46] for a connection of systemic risk between banks and insurers; an excellent review on systemic risk is given by [106].

Extreme co-movement usually requires measuring tail dependence of several variables. Tail dependence has been studied in the context of multivariate extreme value theory for decades. Since such a measure focuses on a far tail region of the underlying distribution, statistical inference is quite challenging due to the lack of observations. Therefore, it is always desirable to find a better measure or some competitive measures and to have an efficient inference procedure.

Suppose (X, Y) is a random vector with joint distribution F and continuous marginal distributions F_1 and F_2. Define $U = 1 - F_1(X)$ and $V = 1 - F_2(Y)$, then the distribution of (U, V) is a survival copula given by

$$C(u, v) = \mathbb{P}(1 - F_1(X) \le u, 1 - F_2(Y) \le v). \tag{5.1}$$

In order to predict an extreme co-movement of financial market, it is useful to investigate the behavior of the so-called tail copula defined as $\lim_{t\to 0} t^{-1}C(tu, tv)$, which can be employed to extrapolate data into a far tail region; see [107] for an overview. When the limit is not identically zero (i.e., asymptotic dependence), one can predict rare events via estimating this limiting function. On the other hand, if the limit is identically zero (i.e., asymptotic independence), then some additional conditions are needed for predicting extreme events. To effectively distinguishing these two cases,[53, 54] introduced the so-called coefficient of tail dependence $\eta \in (0, 1]$ by assuming that $C(t, t) = t^{1/\eta}s(t)$, where $s(t)$ is a slowly varying function, i.e., $\lim_{t\to 0} s(tx)/s(t) = 1$ for all $x > 0$. Therefore, η and the limit of $s(t)$ can be used to distinguish asymptotic dependence (i.e., $\eta = 1 \& \lim_{t\to 0} s(t) > 0$) and asymptotic independence (i.e., $\eta < 1$ or $\eta = 1 \& \lim_{t\to 0} s(t) = 0$). Statistical inference for η is available in [108], [109], [110], and [85].

Although copula gives a complete description of dependence among variables, having some summary measures for dependence is useful in practice. Some commonly used ones include correlation coefficient, Spearman's rho and Kendall's tau. Similarly, tail copula determines the tail dependence completely, but the coefficient of tail dependence η gives a useful measure of tail dependence. Since Kendall's tau is invariant to marginals and has been popular in risk management, one may wonder whether Kendall's tau can be modified to give a simple and effective measure of tail dependence as well. Recently, when the survival copula $C(u,v)$ is a bivariate regular variation, i.e., $H(u,v)=\lim_{t\to 0} C(tu,tv)/C(t,t)$ exists and is finite for $u,v\geq 0$, Asimit et al. (2015) investigated the limit of the conditional Kendall's tau (i.e., $\theta=\lim_{u\to 0}\mathbb{E}\{sgn((U_1-U_2)(V_1-V_2))|\max(U_1,U_2,V_1,V_2)\leq u\}$), found that $\theta=4\int_0^1\int_0^1 H(x,y)dH(x,y)-1$ and showed that θ is positive for a subclass of asymptotic dependence such as elliptical tail copulas and nonpositive for a subclass of asymptotic independence such as normal copulas. Due to its ease of implementation, elliptical copulas and elliptical tail copulas have been employed in risk management; see [90]. The study of tails of mixture of elliptical copulas is available in [111]. A new method for constructing copulas with tail dependence is given by [112]. Since the above measure θ involves the function H rather than some particular values of H as in η, one may expect that θ could be more effective statistically than η in distinguishing asymptotic behavior and measuring tail dependence.

For interval estimation of θ, one can estimate the complicated asymptotic variance of the proposed nonparametric estimator in [104]. In order to avoid estimating the asymptotic variance, a naive bootstrap method can be employed to construct a confidence interval, which generally performs badly in finite sample. Alternatively empirical likelihood methods have been proved to be quite effective in interval estimation and hypothesis test, which requires no estimation for any additional quantities. We refer to [1] for an overview on empirical likelihood methods. In this chapter we investigate the possibility of employing an empirical likelihood method to construct a confidence interval for the limit of the conditional Kendall's tau.

We organize this paper as follows. Section 5.2 presents the new methodology and theoretical results. A simulation study and real data analysis on Danish fire losses are given in Section 5.3. All proofs are put in Section 5.4.

5.2 Methodology and Theoretical Results

Throughout we assume observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ are independent and identically distributed with distribution function F and continuous marginals F_1 and F_2. For the study of asymptotic tail behavior of F, [104] considered the limit of the conditional Kendall's tau, i.e., $\theta = \lim_{u \to 0} \mathbb{E}\{sgn((U_1 - U_2)(V_1 - V_2)) | \max(U_1, U_2, V_1, V_2) \le u\}$. A simple nonparametric estimator for θ is to replace the conditional expectation by its sample conditional mean, which leads to

$$\hat{\theta}(k) = \frac{\sum_{1 \le i < j \le n} sgn((\hat{U}_i - \hat{U}_j)(\hat{V}_i - \hat{V}_j)) I(\max(\hat{U}_i, \hat{U}_j, \hat{V}_i, \hat{V}_j) \le k/n)}{\sum_{1 \le i < j \le n} I(\max(\hat{U}_i, \hat{U}_j, \hat{V}_i, \hat{V}_j) \le k/n)},$$

where $\hat{U}_i = 1 - \hat{F}_1(X_i)$, $\hat{V}_i = 1 - \hat{F}_2(Y_i)$, $\hat{F}_1(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$, $\hat{F}_2(y) = n^{-1} \sum_{i=1}^n I(Y_i \le y)$, $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Under some conditions, [104] derived the asymptotic limit of $\hat{\theta}(k)$, which has a complicated asymptotic variance. Here we investigate the possibility of employing an empirical likelihood method to construct a confidence interval without estimating the asymptotic variance explicitly. By noting that $\hat{\theta}(k)$ is a solution to the following equation

$$\sum_{1 \le i < j \le n} \{sgn((\hat{U}_i - \hat{U}_j)(\hat{V}_i - \hat{V}_j)) - \theta\} I(\max(\hat{U}_i, \hat{U}_j, \hat{V}_i, \hat{V}_j) \le k/n) = 0,$$

one may employ the empirical likelihood method based on estimating equations in [2] to the above equation. Unfortunately such a direct application fails to achieve a chi-squared limit due to the involved U-statistic and the plug-in estimators for $U_i's$ and $V_i's$. Recently a so-called jackknife empirical likelihood method is proposed by [60] to construct confidence intervals for non-linear functions including U-statistics. However, due to the involved indicator function, a direct application of the jackknife empirical likelihood function fails again to have the Wilks theorem. In order to catch the contribution made by the plug-in empirical distributions, we propose to employ the smooth jackknife empirical likelihood method proposed by [58] for constructing confidence intervals for a tail copula.

More specifically, for $l_1, l_2 = 1, \cdots, n$, define

$$\begin{cases} \hat{F}_1^{(l_1)}(x) &= \frac{1}{n}\sum_{j=1, j\neq l_1}^n I(X_j \le x), \quad \hat{U}_{l_2}^{(l_1)} = 1 - \hat{F}_1^{(l_1)}(X_{l_2}), \\ \hat{F}_2^{(l_1)}(x) &= \frac{1}{n}\sum_{j=1, j\neq l_1}^n I(Y_j \le x), \quad \hat{V}_{l_2}^{(l_1)} = 1 - \hat{F}_2^{(l_1)}(Y_{l_2}), \\ \hat{T}_n(\theta) &= \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\{sgn\left((\hat{U}_i - \hat{U}_j)(\hat{V}_i - \hat{V}_j)\right) - \theta\}\times \\ & \quad G(\frac{1-\frac{n}{k}\hat{U}_i}{h})G(\frac{1-\frac{n}{k}\hat{V}_i}{h})G(\frac{1-\frac{n}{k}\hat{U}_j}{h})G(\frac{1-\frac{n}{k}\hat{V}_j}{h}), \\ \hat{T}_n^{(l_1)}(\theta) &= \frac{2}{(n-1)(n-2)}\sum_{1\le i<j\le n,\, i,j\neq l_1}\{sgn\left((\hat{U}_i^{(l_1)} - \hat{U}_j^{(l_1)})(\hat{V}_i^{(l_1)} - \hat{V}_j^{(l_1)})\right) - \theta\}\times \\ & \quad G(\frac{1-\frac{n}{k}\hat{U}_i^{(l_1)}}{h})G(\frac{1-\frac{n}{k}\hat{V}_i^{(l_1)}}{h})G(\frac{1-\frac{n}{k}\hat{U}_j^{(l_1)}}{h})G(\frac{1-\frac{n}{k}\hat{V}_j^{(l_1)}}{h}), \end{cases}$$

where $G(x) = \int_{-\infty}^x g(y)\,dy$ and g is a symmetric smooth density function with support $[-1, 1]$ and $h = h(n) > 0$ is a bandwidth. Therefore a jackknife sample is defined as

$$\hat{Z}_i(\theta) = n\hat{T}_n(\theta) - (n-1)\hat{T}_n^{(i)}(\theta) \quad \text{for} \quad i = 1, \cdots, n.$$

Note that, in order to take care of the contributions from $\hat{U}_i's$ and $\hat{V}_i's$ in proving Wilks theorem, we do not use $G(\frac{1-\frac{n}{k}\max\{\hat{U}_i,\hat{V}_i,\hat{U}_j,\hat{V}_j\}}{h})$ instead of the product of $G's$ in the above definition of $\hat{T}_n(\theta)$. Based on this jackknife sample, a smooth jackknife empirical likelihood function for θ is obtained as

$$L(\theta) = \max\left\{\prod_{i=1}^n (np_i):\ p_1 \ge 0, \cdots, p_n \ge 0\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\hat{Z}_i(\theta) = 0\right\}. \quad (5.2)$$

It follows from the Lagrange multiplier technique that

$$l(\theta) := -2\log L(\theta) = 2\sum_{i=1}^n \log\left\{1 + \lambda\hat{Z}_i(\theta)\right\}, \quad (5.3)$$

where $\lambda = \lambda(\theta)$ satisfies

$$\sum_{i=1}^n \frac{\hat{Z}_i(\theta)}{1 + \lambda\hat{Z}_i(\theta)} = 0.$$

In order to show that Wilks theorem holds for the above smooth jackknife empirical likelihood method, we need some regularity conditions. As usual in extreme value theory, we need a second order regular variation to control the bias in $\hat{\theta}(k)$.

- A1) There exist a regular variation $A(t) \to 0$ as $t \to 0$ with index $\bar{\rho} \geq 0$, functions $Q(u, v)$ and $q(u, v)$ such that

$$\lim_{t\to 0} \frac{\frac{C(tu,tv)}{C(t,t)} - H(u,v)}{A(t)} = Q(u,v) \quad \text{and} \quad \lim_{t\to 0} \frac{\frac{t^2 C_{12}(tu,tv)}{C(t,t)} - H_{12}(u,v)}{A(t)} = q(u,v)$$

  for all $(u, v) \in [0, 1]^2$ and uniformly on $\{(u, v) : u^2 + v^2 = 1\}$, where $H_{12}(u, v) = \frac{\partial^2}{\partial u \partial v} H(u, v)$ and $C_{12}(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$;

- A2) $k \to \infty$, $\frac{n}{k} C(\frac{k}{n}, \frac{k}{n}) \to c_0 \in [0, 1]$, $nC(\frac{k}{n}, \frac{k}{n}) \to \infty$, $\{nC(\frac{k}{n}, \frac{k}{n})\}^{1/2} A(\frac{k}{n}) \to 0$ as $n \to \infty$;

- A3) $h \to 0$, $kh \to \infty$, $nC(\frac{k}{n}, \frac{k}{n}) h^2 \to \infty$, $nC(\frac{k}{n}, \frac{k}{n}) h^4 \to 0$, $\frac{\sqrt{nC(\frac{k}{n}, \frac{k}{n})}}{kh} \to 0$ as $n \to \infty$.

Theorem 5.2.1. *Under conditions A1)–A3), $l(\theta_0)$ converges in distribution to a chi-squared limit with one degree of freedom as $n \to \infty$, where θ_0 is the true value of θ, i.e., the true limit of the conditional Kendall's tau.*

Based on the above limit, a confidence interval for θ_0 with level γ is $I_\gamma^{EL} = \{\theta : l(\theta) \leq \chi_{1,\gamma}\}$, where $\chi_{1,\gamma}$ is the γ-th quantile of a chi-squared distribution with one degree of freedom. In the simulation study below, we provide a way to choose h, which is less important than choosing k in general. As usual in extreme value theory, it is always challenging to choose k. We plan to investigate the issue of data-driven methods for choosing k in the future.

5.3 Simulation Study and Data Analysis

First we examine the finite sample behavior of the proposed jackknife empirical likelihood method in terms of coverage accuracy and compare it with the normal approximation method.

Draw $1,000$ random samples of size $n = 1,000$ from a normal copula with correlation ρ and the elliptical random variable RAU, where $R > 0$ is a random variable with distribution $P(R > x) = x^{-\alpha}$ for some $\alpha > 0$, A is a deterministic 2×2 matrix with $AA^T = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, U is uniformly distributed on $\{z = (z_1, z_2)^T : z^T z = 1\}$ and independent of R. Hence the true θ for normal copula is zero. A formula for computing the true θ for the above elliptical distribution is given in [104].

Motivated by the choice of bandwidth in smoothing distribution function estimation, we choose $h = \delta\{\sum_{i=1}^n I(\hat{U}_i \le \frac{k}{n}, \hat{V}_i \le \frac{k}{n})\}^{-1/3}$ with $\delta = 0.5, 1, 1.5$. We employ the kernel $g(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \le 1)$ and consider $k = 50, 100, 150, 200$. For computing the confidence interval based on the smooth estimator $\tilde{\theta}$ which solves $\hat{T}_n(\theta) = 0$, we use the bootstrap method by drawing $1,000$ resamples. Denote this interval by I_γ^B.

Coverage probabilities for the bootstrap method (I_γ^B) and the empirical likelihood method (I_γ^{EL}) with level $\gamma = 0.9$ and 0.95 are reported in Table 1, which shows that i) the proposed empirical likelihood method performs better in most cases; ii) both methods are less sensitive to the choice of bandwidth; iii) the proposed empirical likelihood method is more robust with respect to the choice of k for elliptical distributions.

Next, we consider the non-zero losses to building and content in the Danish fire insurance claims; see Figure 1. This data set is available at *www.ma.hw.ac.uk/∼mcneil/*, which comprises 2,167 fire losses over the period 1980 to 1990. Figure 1 shows that there are some huge losses to both content and building, but a few simultaneous large losses to both variables, which may suggest a weak tail dependence. In Table 2 below we report confidence intervals I_γ^B and I_γ^{EL} with level $\gamma = 0.9$ and 0.95. As above, the bootstrap confidence

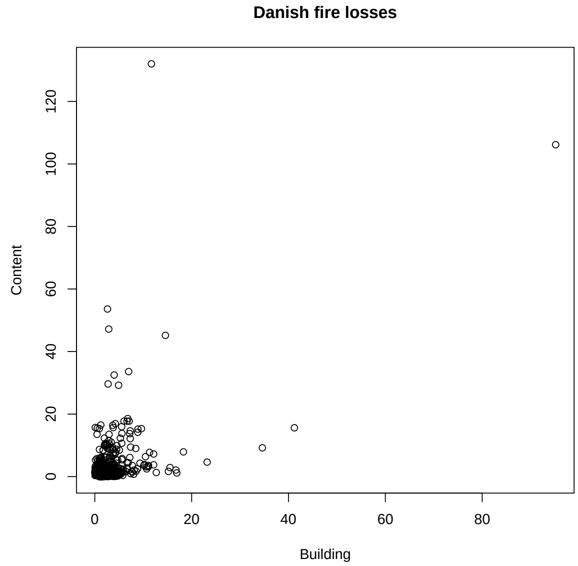


Figure 5.1: Danish fire losses.

interval is based on the smooth estimator $\tilde{\theta}$ and $1,000$ bootstrap resamples, and the same kernel function G and the choice of h are employed. For computing the empirical likelihood based confidence interval, we calculate the empirical likelihood ration function for θ from -0.6 to 0.6 with a step 0.01. Table 2 shows that the empirical likelihood confidence intervals are slightly more skewed to the right than the normal approximation confidence intervals. The results for $k = 90$ and 100 may prefer a positive θ than nonnegative one, which may indicate a weak tail dependence as suggested by Figure 1.

5.4 Proofs

Throughout we define

$$
\begin{aligned}
B_{lm} &= sgn((\hat{U}_l - \hat{U}_m)(\hat{V}_l - \hat{V}_m)) - \theta_0 = sgn((U_l - U_m)(V_l - V_m)) - \theta_0, \\
g_h(x) &= g\left(\frac{1 - \frac{n}{k}x}{h}\right), \quad G_h(x) = G\left(\frac{1 - \frac{n}{k}x}{h}\right), \\
\bar{A}_{lm} &= B_{lm} G_h(U_l) G_h(U_m) G_h(V_l) G_h(V_m), \\
A_{lm} &= B_{lm} G_h(\hat{U}_l) G_h(\hat{U}_m) G_h(\hat{V}_l) G_h(\hat{V}_m), \\
A_{lm}^{(i)} &= B_{lm} G_h(\hat{U}_l^{(i)}) G_h(\hat{U}_m^{(i)}) G_h(\hat{V}_l^{(i)}) G_h(\hat{V}_m^{(i)}).
\end{aligned}
$$

Hence $\hat{T}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{1 \le l < m \le n} A_{lm}$ and $\hat{T}_n^{(i)}(\theta_0) = \frac{2}{(n-1)(n-2)} \sum_{1 \le l < m \le n, l \ne i, m \ne i} A_{lm}^{(i)}$.

Lemma 5.4.1. *Under conditions A1)–A3), we have*

$$
\begin{aligned}
& \frac{n}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \hat{T}_n(\theta_0) \\
= \; & \frac{2}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \sum_{i=1}^n h_0(U_i, V_i) \\
& + \sqrt{k}\{\tfrac{1}{k} \sum_{i=1}^n I(U_i \le \tfrac{k}{n}) - 1\} \sqrt{c_0} \{4 \int_0^1 H_{12}(1, v) H(1, v)\, dv - 2(1 + \theta_0) H_1(1, 1)\} \\
& + \sqrt{k}\{\tfrac{1}{k} \sum_{i=1}^n I(V_i \le \tfrac{k}{n}) - 1\} \sqrt{c_0} \{4 \int_0^1 H_{12}(u, 1) H(u, 1)\, du - 2(1 + \theta_0) H_2(1, 1)\} + o_p(1) \\
:= \; & W_{n1} + W_{n2} + W_{n3} + o_p(1),
\end{aligned}
\tag{5.4}
$$

and

$$
\frac{n}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \hat{T}_n(\theta_0) \xrightarrow{d} N(0, \sigma^2), \tag{5.5}
$$

where

$$
\begin{aligned}
h_0(u, v) &= \{4C(u, v) - 2C(u, \tfrac{k}{n}) - 2C(\tfrac{k}{n}, v) + (1 - \theta_0) C(\tfrac{k}{n}, \tfrac{k}{n})\} I(\max(u, v) \le \tfrac{k}{n}), \\
\sigma^2 &= 4\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2c_0 \sigma_2 \sigma_3,
\end{aligned}
$$

$$\begin{cases} \sigma_1^2 = \int_0^1 \int_0^1 \{4H(x,y) - 2H(x,1) - 2H(1,y) + (1-\theta_0)\}^2 \, dH(x,y), \\ \sigma_2 = \sqrt{c_0}\{4\int_0^1 H_{12}(1,v)H(1,v)\, dv - 2(1+\theta_0)H_1(1,1)\}, \\ \sigma_3 = \sqrt{c_0}\{4\int_0^1 H_{12}(u,1)H(u,1)\, du - 2(1+\theta_0)H_2(1,1)\}. \end{cases}$$

Proof. Put

$$\begin{cases} \theta_n = E\{(sgn\,((U_1-U_2)(V_1-V_2)) - \theta_0)G_h(U_1)G_h(V_1)G_h(U_2)G_h(V_2)\}, \\ \tilde{h}(u_1,v_1,u_2,v_2) = \{sgn\,((u_1-u_2)(v_1-v_2)) - \theta_0\}G_h(u_1)G_h(v_1)G_h(u_2)G_h(v_2) - \theta_n, \\ \tilde{h}_1(u_1,v_1) = E\{(sgn\,((u_1-U_2)(v_1-V_2)) - \theta_0)G_h(u_1)G_h(v_1)G_h(U_2)G_h(V_2)\} - \theta_n, \\ S_{1n} = \sum_{i=1}^n \tilde{h}_1(U_i,V_i), \\ S_{2n} = \sum_{1\le l<m\le n}\{\tilde{h}(U_l,V_l,U_m,V_m) - \tilde{h}_1(U_l,V_l) - \tilde{h}_1(U_m,V_m)\}, \\ \tilde{T}_n(\theta_0) = \frac{2}{n(n-1)}\sum_{1\le l<m\le n}\{sgn\,((U_l-U_m)(V_l-V_m)) - \theta_0\}G_h(U_l)G_h(V_l)G_h(U_m)G_h(V_m) \\ \qquad\qquad -\theta_n. \end{cases}$$

Since $\tilde{T}_n(\theta_0)$ is a U-statistic, from the Hoeffding decomposition (see [113] or Lemma A from Page 178 of [114]), we have

$$\tilde{T}_n(\theta) = \frac{2}{n}S_{1n} + \frac{2}{n(n-1)}S_{2n}. \tag{5.6}$$

It is straightforward to check that

$$\begin{array}{ll} & \frac{\theta_n}{C^2(\frac{k}{n},\frac{k}{n})} \\ = & \frac{1}{C^2(\frac{k}{n},\frac{k}{n})}\int_{[0,1]^4}\{sgn((u_1-u_2)(v_1-v_2)) - \theta_0\}G_h(u_1)G_h(v_1)G_h(u_2)G_h(v_2)dC(u_1,v_1)dC(u_2,v_2) \\ = & \frac{1}{C^2(\frac{k}{n},\frac{k}{n})}\int_{[-1,1]^4}\prod_{i=1}^4 g(t_i)\int_{\prod_{i=1}^4[0,(1-t_ih)\frac{k}{n}]^4}\{sgn((u_1-u_2)(v_1-v_2)) - \theta_0\} \\ & dC(u_1,v_1)dC(u_2,v_2)\prod_{i=1}^4 dt_i \\ \to & \int_{[-1,1]^4}\prod_{i=1}^4 g(t_i)\int_{[0,1]^4} sgn((x_1-x_2)(y_1-y_2)-\theta_0)dH(x_1,y_1)dH(x_2,y_2)\prod_{i=1}^4 dt_i \\ = & 0, \end{array} \tag{5.7}$$

and

$$
\begin{aligned}
& \tilde{h}_1(u_1, v_1) \\
= \; & 4E\left\{I(U_2 < u_1, V_2 < v_1)G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)G\left(\frac{1-\frac{n}{k}U_2}{h}\right)G\left(\frac{1-\frac{n}{k}V_2}{h}\right)\right\} \\
& -2E\left\{I(U_2 < u_1)G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)G\left(\frac{1-\frac{n}{k}U_2}{h}\right)G\left(\frac{1-\frac{n}{k}V_2}{h}\right)\right\} \\
& -2E\left\{I(V_2 < v_1)G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)G\left(\frac{1-\frac{n}{k}U_2}{h}\right)G\left(\frac{1-\frac{n}{k}V_2}{h}\right)\right\} \\
& +(1-\theta_0)E\left\{G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)G\left(\frac{1-\frac{n}{k}U_2}{h}\right)G\left(\frac{1-\frac{n}{k}V_2}{h}\right)\right\} - \theta_n \\
= \; & 4G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)C(u_1, v_1) - 2G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)\int_{-1}^{1} g(t)C(u_1, (1-th)\tfrac{k}{n})dt \\
& -2G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)\int_{-1}^{1} g(t)C((1-th)\tfrac{k}{n}, v_1)dt \\
& +(1-\theta_0)G\left(\frac{1-\frac{n}{k}u_1}{h}\right)G\left(\frac{1-\frac{n}{k}v_1}{h}\right)\int_{[-1,1]^2} g(t_1)g(t_2)C((1-t_1h)\tfrac{k}{n}, (1-t_2h)\tfrac{k}{n})dt_1dt_2 - \theta_n \\
= \; & (1+O(h))\{4C(u_1, v_1)I(\max(u_1, v_2) \le \tfrac{k}{n}) - 2C(u_1, \tfrac{k}{n})I(\max(u_1, v_1) \le \tfrac{k}{n}) \\
& -2C(\tfrac{k}{n}, v_1)I(\max(u_1, v_1) \le \tfrac{k}{n}) + (1-\theta_0)C(\tfrac{k}{n}, \tfrac{k}{n})I(\max(u_1, v_1) \le \tfrac{k}{n})\} - \theta_n.
\end{aligned}
\tag{5.8}
$$

Similar to the proof of (5.7), we have

$$
\frac{E\tilde{h}_1^2(U_1, V_1)}{C^3(\frac{k}{n}, \frac{k}{n})} \to \int_0^1\int_0^1 \{4H(x,y) - 2H(x,1) - 2H(1,y) + (1-\theta_0)\}^2\, dH(x,y) \tag{5.9}
$$

and

$$
\frac{E\tilde{h}^2(U_1, V_1, U_2, V_2)}{C^2(\frac{k}{n}, \frac{k}{n})} \to 1 - \theta_0^2. \tag{5.10}
$$

It is derived from page 178 and 184 of Serfling (1980) that

$$
ES_{2n}^2 = \binom{n}{2} Eg_2^2, \tag{5.11}
$$

where $g_2 = \tilde{h}(U_1, V_1, U_2, V_2) - \tilde{h}_1(U_1, V_1) - \tilde{h}_1(U_2, V_2)$. From (5.9) and (5.10), we have $\frac{Eg_2^2}{nC^3(\frac{k}{n}, \frac{k}{n})} \to 0$, which implies

$$
\frac{S_{2n}}{\sqrt{n^3C^3(\frac{k}{n}, \frac{k}{n})}} = o_p(1) \tag{5.12}
$$

by using (5.11).

Note that

$$
\begin{aligned}
& \frac{1}{C^2(\frac{k}{n},\frac{k}{n})} E G_h(U_1)G_h(V_1)G_h(U_2)G_h(V_2) \\
= & \left\{ \frac{1}{C(\frac{k}{n},\frac{k}{n})} \int_{[-1,1]^2} \int_0^{\frac{k}{n}(1-t_1h)} \int_0^{\frac{k}{n}(1-t_2h)} g(t_1)g(t_2) dC(u_1,v_1) dt_1 dt_2 \right\}^2 \\
= & \left\{ \frac{1}{C(\frac{k}{n},\frac{k}{n})} \int_{[-1,1]^2} \int_0^{(1-t_1h)} \int_0^{(1-t_2h)} g(t_1)g(t_2) dC(\tfrac{k}{n}x,\tfrac{k}{n}y) dt_1 dt_2 \right\}^2.
\end{aligned}
$$

Assumption A1) leads to

$$
\begin{aligned}
& \frac{1}{A(\frac{k}{n})} \Big\{ \frac{1}{C(\frac{k}{n},\frac{k}{n})} \int_{[-1,1]^2} \int_0^{(1-t_1h)} \int_0^{(1-t_2h)} g(t_1)g(t_2) dC(\tfrac{k}{n}x,\tfrac{k}{n}y) dt_1 dt_2 \\
& \qquad - \int_{[-1,1]^2} g(t_1)g(t_2) H(1-t_1h, 1-t_2h) dt_1 dt_2 \Big\} \\
= & \int_{[-1,1]^2} g(t_1)g(t_2) \int_0^{(1-t_1h)} \int_0^{(1-t_2h)} d\left(\frac{\frac{C(\frac{k}{n}x,\frac{k}{n}y)}{C(\frac{k}{n},\frac{k}{n})} - H(x,y)}{A(\frac{k}{n})} \right) dt_1 dt_2 \\
\to & \; Q(1,1),
\end{aligned}
$$

and by Taylor expansion and the symmetry of g, we have

$$
\int_{[-1,1]^2} g(t_1)g(t_2) H(1-t_1h, 1-t_2h) dt_1 dt_2 = H(1,1) + O(h^2).
$$

Thus,

$$
\begin{aligned}
& \frac{1}{C(\frac{k}{n},\frac{k}{n})} \int_{[-1,1]^2} \int_0^{(1-t_1h)} \int_0^{(1-t_2h)} g(t_1)g(t_2) dC(\tfrac{k}{n}x,\tfrac{k}{n}y) dt_1 dt_2 \\
= & \; 1 + O(h^2) + O(A(\tfrac{k}{n})),
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \frac{2\theta_0}{n(n-1)C^2(\frac{k}{n},\frac{k}{n})} \sum_{1\le l<m\le n} G_h(U_1)G_h(V_1)G_h(U_2)G_h(V_2) \\
= & \; \theta_0\{1 + O_p(h^2) + O_p(A(\tfrac{k}{n}))\}^2 \qquad (5.13) \\
= & \; \theta_0 + O_p(h^2) + O_p(A(\tfrac{k}{n})).
\end{aligned}
$$

Similar to the proof of (5.13), we can show that

$$\frac{\theta_n}{C^2(\frac{k}{n},\frac{k}{n})} = O(h^2) + O(A(\tfrac{k}{n})). \tag{5.14}$$

Therefore it follows from (5.6), (5.8), (5.12) and (5.14) that

$$\begin{aligned} \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\tilde{T}_n(\theta_0) &= \frac{2}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\sum_{i=1}^n \tilde{h}_1(U_i,V_i) + \frac{2}{(n-1)\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}S_{2n} \\ &= \frac{2}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\sum_{i=1}^n h_0(U_i,V_i) + o_p(1). \end{aligned} \tag{5.15}$$

Denote $G_{n1}(x) = \frac{1}{n}\sum_{i=1}^n I(U_i \le x)$ and $G_{n2}(y) = \frac{1}{n}\sum_{i=1}^n I(V_i \le y)$. Then $\hat{U}_i = \hat{G}_{n1}(U_i)$ and $\hat{V}_i = G_{n2}(V_i)$ for $i = 1, \cdots, n$. By Taylor expansion, we have

$$\begin{aligned} & A_{lm} - \bar{A}_{lm} \\ = \ & B_{lm}g_h(U_l)G_h(U_m)G_h(V_l)G_h(V_m)\tfrac{n}{kh}(G_{n1}(U_l) - U_l) \\ & +B_{lm}G_h(U_l)g_h(U_m)G_h(V_l)G_h(V_m)\tfrac{n}{kh}(G_{n1}(U_m) - U_m) \\ & +B_{lm}G_h(U_l)G_h(U_m)g_h(V_l)G_h(V_m)\tfrac{n}{kh}(G_{n2}(V_l) - V_l) \\ & +B_{lm}G_h(U_l)G_h(U_m)G_h(V_l)g_h(V_m)\tfrac{n}{kh}(G_{n2}(V_m) - V_m) \\ & +\tfrac{1}{2}B_{lm}g_h'(\xi_1(i,l))G_h(U_m)G_h(V_l)G_h(V_m)[\tfrac{n}{kh}(G_{n1}(U_l) - U_l)]^2 \\ & +\tfrac{1}{2}B_{lm}G_h(U_l)g_h'(\xi_1(i,m))G_h(V_l)G_h(V_m)[\tfrac{n}{kh}(G_{n1}(U_m) - U_m)]^2 \\ & +\tfrac{1}{2}B_{lm}G_h(U_l)G_h(U_m)g_h'(\xi_2(i,l))G_h(V_m)[\tfrac{n}{kh}(G_{n2}(V_l) - V_l)]^2 \\ & +\tfrac{1}{2}B_{lm}G_h(U_l)G_h(U_m)G_h(V_l)g_h'(\xi_2(i,m))[\tfrac{n}{kh}(G_{n2}(V_m) - V_m)]^2 \\ & +B_{lm}g_h(\xi_1(i,l))g_h(\xi_1(i,m))G_h(V_l)G_h(V_m)(\tfrac{n}{kh})^2(G_{n1}(U_l) - U_l)(G_{n1}(U_m) - U_m) \\ & +B_{lm}g_h(\xi_1(i,l))g_h(\xi_2(i,l))G_h(U_m)G_h(V_m)(\tfrac{n}{kh})^2(G_{n1}(U_l) - U_l)(G_{n2}(V_l) - V_l) \\ & +B_{lm}g_h(\xi_1(i,l))g_h(\xi_2(i,m))G_h(U_m)G_h(V_l)(\tfrac{n}{kh})^2(G_{n1}(U_l) - U_l)(G_{n2}(V_m) - V_m) \\ & +B_{lm}g_h(\xi_1(i,m))g_h(\xi_2(i,l))G_h(U_l)G_h(V_m)(\tfrac{n}{kh})^2(G_{n1}(U_m) - U_m)(G_{n2}(V_l) - V_l) \\ & +B_{lm}g_h(\xi_1(i,m))g_h(\xi_2(i,m))G_h(U_l)G_h(V_l)(\tfrac{n}{kh})^2(G_{n1}(U_m) - U_m)(G_{n2}(V_m) - V_m) \\ & +B_{lm}g_h(\xi_2(i,l))g_h(\xi_2(i,m))G_h(U_l)G_h(U_m)(\tfrac{n}{kh})^2(G_{n2}(V_l) - V_l)(G_{n2}(V_m) - V_m) \\ := \ & DA1_{lm} + DA2_{lm} + DA3_{lm} + DA4_{lm} + \sum_{k=1}^{10} DBk_{lm}, \end{aligned} \tag{5.16}$$

where $\xi_1(i,k), k=l,m$ is between $G_{n1}(U_k)$ and U_k, and $\xi_2(i,k)$ is between $G_{n2}(V_k)$ and V_k. It is straightforward to show that

$$\begin{aligned}
& \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)}\sum_{1\le l<m\le n} DA1_{lm} \\
= & \sqrt{\frac{n}{k}C(\frac{k}{n},\frac{k}{n})}\frac{1}{n(n-1)C^2(\frac{k}{n},\frac{k}{n})}\sum_{l\neq m} B_{lm}h^{-1}g_h(U_l)G_h(V_l)G_h(U_m)G_h(V_m)\times \\
& \sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(U_i\le U_l)-\frac{n}{k}U_l\} \\
= & \sqrt{c_0}\sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(U_i\le \frac{k}{n})-1\}\times \\
& \int_0^1\int_0^1\int_0^1\{sgn(v_1-v_2)-\theta_0\}H_{12}(1,v_1)H_{12}(u_2,v_2)\,dv_1du_2dv_2+o_p(1) \\
= & \sqrt{c_0}\sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(U_i\le \frac{k}{n})-1\}\{2\int_0^1 H_{12}(1,v)H(1,v)\,dv-(1+\theta_0)H_1(1,1)\}+o_p(1),
\end{aligned} \tag{5.17}$$

$$\begin{aligned}
& \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)}\sum_{1\le l<m\le n} DA2_{lm} \\
= & \sqrt{c_0}\sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(U_i\le \frac{k}{n})-1\}\{2\int_0^1 H_{12}(1,v)H(1,v)\,dv-(1+\theta_0)H_1(1,1)\}+o_p(1),
\end{aligned} \tag{5.18}$$

$$\begin{aligned}
& \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)}\sum_{1\le l<m\le n} DA3_{lm} \\
= & \sqrt{c_0}\sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(V_i\le \frac{k}{n})-1\}\{2\int_0^1 H_{12}(u,1)H(u,1)\,du-(1+\theta_0)H_2(1,1)\}+o_p(1),
\end{aligned} \tag{5.19}$$

$$\begin{aligned}
& \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)}\sum_{1\le l<m\le n} DA4_{lm} \\
= & \sqrt{c_0}\sqrt{k}\{\frac{1}{k}\sum_{i=1}^n I(V_i\le \frac{k}{n})-1\}\{2\int_0^1 H_{12}(u,1)H(u,1)\,du-(1+\theta_0)H_2(1,1)\}+o_p(1).
\end{aligned} \tag{5.20}$$

Set $C_0=\max(1,\max_{x\in[-1,1]}g(x),\max_{x\in[-1,1]}|g'(x)|,1-\theta_0,1+\theta_0,|\theta_0|)$. Since g is a smooth density function with support $[-1,1]$, we have

$$\begin{array}{rl} & \sum_{l=1}^{n} |g_h'(\xi_1(i,l))| \\ \leq & C_0 \sum_{l=1}^{n} I\left(-1 \leq \frac{1-\frac{n}{k}\hat{U}_l}{h} \leq 1\right) \\ \leq & C_0 \sum_{l=1}^{n} I\left(-1 \leq \frac{1-\frac{n}{k}\hat{U}_l}{h} \leq 1\right) \\ \leq & C_0 \sum_{l=1}^{n} I\left(\hat{F}_1^{-}(1-\frac{k}{n}(1+h)) \leq X_l < \hat{F}_1^{-}(1-\frac{k}{n}(1-h)+\frac{1}{n})\right) \\ \leq & C_0 n \left(\hat{F}_1(\hat{F}_1^{-}(1-\frac{k}{n}(1-h)+\frac{1}{n})^{-}) - \hat{F}_1(\hat{F}_1^{-}(1-\frac{k}{n}(1+h))) + \frac{1}{n}\right) \\ \leq & C_0 n \left(1-\frac{k}{n}(1-h)+\frac{1}{n}-1+\frac{k}{n}(1+h)+\frac{1}{n}\right) \\ \leq & 2C_0(1+kh). \end{array} \tag{5.21}$$

By (5.21), we can show that

$$\begin{array}{rl} & \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \frac{2}{n(n-1)} \sum_{1\leq l<m\leq n} |DBj_{lm}| \\ = & O_p(\frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \frac{1}{n^2} khnC(\frac{k}{n},\frac{k}{n}) \frac{1}{kh^2}) = O_p(\frac{1}{h\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1) \end{array} \tag{5.22}$$

for $j = 1, 2, 3, 4$, and

$$\begin{array}{rl} & \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \frac{2}{n(n-1)} \sum_{1\leq l<m\leq n} DBj_{lm} \\ = & O_p(\frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} C^2(\frac{k}{n},\frac{k}{n}) \frac{1}{k}) = O_p(\frac{c_0}{\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1) \end{array} \tag{5.23}$$

for $j = 5, 6, 7, 8, 9, 10$. Hence (5.4) follows from (5.15)–(5.20), (5.22) and (5.23).

Further we have

$$\left\{ \begin{array}{l} E\left\{\frac{h_0(U_i,V_i)}{\sqrt{Eh_0^2(U_1,V_1)}} \frac{I(U_i\leq\frac{k}{n})-\frac{k}{n}}{\sqrt{k/n}}\right\} = \frac{\theta_n-\frac{k}{n}\theta_n}{\sigma_1 C(\frac{k}{n},\frac{k}{n})} \sqrt{\frac{n}{k}C(\frac{k}{n},\frac{k}{n})}\{1+o(1)\} \to 0, \\ E\left\{\frac{h_0(U_i,V_i)}{\sqrt{Eh_0^2(U_1,V_1)}} \frac{I(V_i\leq\frac{k}{n})-\frac{k}{n}}{\sqrt{k/n}}\right\} = \frac{\theta_n-\frac{k}{n}\theta_n}{\sigma_1 C(\frac{k}{n},\frac{k}{n})} \sqrt{\frac{n}{k}C(\frac{k}{n},\frac{k}{n})}\{1+o(1)\} \to 0, \\ E\left\{\frac{I(U_i\leq\frac{k}{n})-\frac{k}{n}}{\sqrt{k/n}} \frac{I(U_i\leq\frac{k}{n})-\frac{k}{n}}{\sqrt{k/n}}\right\} = \frac{C(\frac{k}{n},\frac{k}{n})-(\frac{k}{n})^2}{k/n} \to c_0. \end{array} \right.$$

Consequently, it follows from the Cramér-device that

$$(W_{n1}, W_{n2}, W_{n3})^T = \left(\frac{2\sigma_1}{\sqrt{n}} \sum_{i=1}^{n} \frac{h_0(U_i, V_i)}{\sqrt{Eh_0^2(U_1, V_1)}}, W_{n2}, W_{n3} \right)^T + o_p(1) \xrightarrow{d} N(0, \Sigma)$$

as $n \to \infty$, where

$$\Sigma = \begin{pmatrix} 4\sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & c_0\sigma_2\sigma_3 \\ 0 & c_0\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix},$$

which implies (5.5).

□

Lemma 5.4.2. *Under condition A1) –A3), we have*

$$\frac{1}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \sum_{i=1}^{n} \hat{Z}_i(\theta_0) \xrightarrow{d} N(0, \sigma^2) \tag{5.24}$$

as $n \to \infty$, where $\sigma^2 = \lim_{n\to\infty} E(W_{n1} + W_{n2} + W_{n3})^2$ is given in (5.5).

Proof. According to the definition, we have

$$\frac{1}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \sum_{i=1}^{n} \hat{Z}_i(\theta_0) = \frac{n}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \hat{T}_n(\theta_0) + \frac{n-1}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \sum_{i=1}^{n} [\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0)].$$

By Lemma 5.4.1, to obtain (5.24), it is sufficient to prove

$$\frac{n-1}{\sqrt{nC^3(\frac{k}{n}, \frac{k}{n})}} \sum_{i=1}^{n} [\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0)] = o_p(1). \tag{5.25}$$

Note that $A_{lm}^{(i)} = A_{ml}^{(i)}$, which allows us to write

$$\begin{aligned} \hat{T}_n^{(i)}(\theta_0) &= \tfrac{2}{(n-1)(n-2)} [\textstyle\sum_{1 \le l < m \le n} A_{lm}^{(i)} - \sum_{m > i} A_{im}^{(i)} - \sum_{l < i} A_{li}^{(i)}] \\ &= \tfrac{2}{(n-1)(n-2)} [\textstyle\sum_{1 \le l < m \le n} A_{lm}^{(i)} - \sum_{l=1}^{n} A_{li}^{(i)} + A_{ii}^{(i)}]. \end{aligned}$$

Thus,

$$
\begin{aligned}
\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0) &= \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{1\le l<m\le n} [A_{lm} - A_{lm}^{(i)}] \\
&\quad - \tfrac{2}{(n-1)(n-2)} [\tfrac{1}{n} \textstyle\sum_{l=1}^n \sum_{j=1}^n A_{lj} - \sum_{l=1}^n A_{li}^{(i)}] \\
&\quad + \tfrac{2}{(n-1)(n-2)} [\tfrac{1}{n} \textstyle\sum_{j=1}^n A_{jj} - A_{ii}^{(i)}] \\
&:= D_{1i} - D_{2i} + D_{3i}
\end{aligned}
\tag{5.26}
$$

and

$$
\begin{aligned}
\textstyle\sum_{i=1}^n [\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0)] &= \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} [A_{lm} - A_{lm}^{(i)}] \\
&\quad - \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n [\tfrac{1}{n} \sum_{l=1}^n \sum_{j=1}^n A_{lj} - \sum_{l=1}^n A_{li}^{(i)}] \\
&\quad + \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n [\tfrac{1}{n} \sum_{j=1}^n A_{jj} - A_{ii}^{(i)}] \\
&= \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} [A_{lm} - A_{lm}^{(i)}] \\
&\quad - \tfrac{2}{(n-1)(n-2)} [\textstyle\sum_{l=1}^n \sum_{i=1}^n A_{li} - \sum_{i=1}^n \sum_{l=1}^n A_{li}^{(i)}] \\
&\quad + \tfrac{2}{(n-1)(n-2)} [\textstyle\sum_{i=1}^n A_{ii} - \sum_{i=1}^n A_{ii}^{(i)}] \\
&= \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} [A_{lm} - A_{lm}^{(i)}] \\
&\quad - \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{l=1}^n [A_{li} - A_{li}^{(i)}] \\
&\quad + \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n [A_{ii} - A_{ii}^{(i)}] \\
&:= \textstyle\sum_{i=1}^n \tilde{D}_{1i} - \sum_{i=1}^n \tilde{D}_{2i} + \sum_{i=1}^n \tilde{D}_{3i}.
\end{aligned}
\tag{5.27}
$$

By Taylor expansion, we have

$$
\begin{aligned}
& A_{lm} - A_{lm}^{(i)} \\
= \; & B_{lm}\{G_h(\hat{U}_l)G_h(\hat{U}_m)G_h(\hat{V}_l)G_h(\hat{V}_m) - G_h(\hat{U}_l^{(i)})G_h(\hat{U}_m^{(i)})G_h(\hat{V}_l^{(i)})G_h(\hat{V}_m^{(i)})\} \\
= \; & B_{lm}g_h(\hat{U}_l)G_h(\hat{U}_m)G_h(\hat{V}_l)G_h(\hat{V}_m)\tfrac{n}{kh}(\hat{U}_l - \hat{U}_l^{(i)}) \\
& +B_{lm}G_h(\hat{U}_l)g_h(\hat{U}_m)G_h(\hat{V}_l)G_h(\hat{V}_m)\tfrac{n}{kh}(\hat{U}_m - \hat{U}_m^{(i)}) \\
& +B_{lm}G_h(\hat{U}_l)G_h(\hat{U}_m)g_h(\hat{V}_l)G_h(\hat{V}_m)\tfrac{n}{kh}(\hat{V}_l - \hat{V}_l^{(i)}) \\
& +B_{lm}G_h(\hat{U}_l)G_h(\hat{U}_m)G_h(\hat{V}_l)g_h(\hat{V}_m)\tfrac{n}{kh}(\hat{V}_m - \hat{V}_m^{(i)}) \\
& +\tfrac{1}{2}B_{lm}g_h'(\xi_1(i,l))G_h(\hat{U}_m)G_h(\hat{V}_l)G_h(\hat{V}_m)[\tfrac{n}{kh}(\hat{U}_l - \hat{U}_l^{(i)})]^2 \\
& +\tfrac{1}{2}B_{lm}G_h(\hat{U}_l)g_h'(\xi_1(i,m))G_h(\hat{V}_l)G_h(\hat{V}_m)[\tfrac{n}{kh}(\hat{U}_m - \hat{U}_m^{(i)})]^2 \\
& +\tfrac{1}{2}B_{lm}G_h(\hat{U}_l)G_h(\hat{U}_m)g_h'(\xi_2(i,l))G_h(\hat{V}_m)[\tfrac{n}{kh}(\hat{V}_l - \hat{V}_l^{(i)})]^2 \\
& +\tfrac{1}{2}B_{lm}G_h(\hat{U}_l)G_h(\hat{U}_m)G_h(\hat{V}_l)g_h'(\hat{\xi}_2(i,m))[\tfrac{n}{kh}(\hat{V}_m - \hat{V}_m^{(i)})]^2 \\
& +B_{lm}g_h(\xi_1(i,l))g_h(\xi_1(i,m))G_h(\hat{V}_l)G_h(\hat{V}_m)(\tfrac{n}{kh})^2(\hat{U}_l - \hat{U}_l^{(i)})(\hat{U}_m - \hat{U}_m^{(i)}) \\
& +B_{lm}g_h(\xi_1(i,l))g_h(\xi_2(i,l))G_h(\hat{U}_m)G_h(\hat{V}_m)(\tfrac{n}{kh})^2(\hat{U}_l - \hat{U}_l^{(i)})(\hat{V}_l - \hat{V}_l^{(i)}) \\
& +B_{lm}g_h(\xi_1(i,l))g_h(\xi_2(i,m))G_h(\hat{U}_m)G_h(\hat{V}_l)(\tfrac{n}{kh})^2(\hat{U}_l - \hat{U}_l^{(i)})(\hat{V}_m - \hat{V}_m^{(i)}) \\
& +B_{lm}g_h(\xi_1(i,m))g_h(\xi_2(i,l))G_h(\hat{U}_l)G_h(\hat{V}_m)(\tfrac{n}{kh})^2(\hat{U}_m - \hat{U}_m^{(i)})(\hat{V}_l - \hat{V}_l^{(i)}) \\
& +B_{lm}g_h(\xi_1(i,m))g_h(\xi_2(i,m))G_h(\hat{U}_l)G_h(\hat{V}_l)(\tfrac{n}{kh})^2(\hat{U}_m - \hat{U}_m^{(i)})(\hat{V}_m - \hat{V}_m^{(i)}) \\
& +B_{lm}g_h(\xi_2(i,l))g_h(\xi_2(i,m))G_h(\hat{U}_l)G_h(\hat{U}_m)(\tfrac{n}{kh})^2(\hat{V}_l - \hat{V}_l^{(i)})(\hat{V}_m - \hat{V}_m^{(i)}) \\
:= \; & DA1_{lm,i} + DA2_{lm,i} + DA3_{lm,i} + DA4_{lm,i} + \textstyle\sum_{k=1}^{10} DBk_{lm,i},
\end{aligned}
\tag{5.28}
$$

where $\xi_1(i,k)$, $k = l, m$, is between $\hat{U}_k$ and $\hat{U}_k^{(i)}$, and $\xi_2(i,k)$ is between $\hat{V}_k$ and $\hat{V}_k^{(i)}$.

Since

$$\hat{U}_l - \hat{U}_l^{(i)} = \frac{1}{n-1}I(U_i < U_l) - \frac{1}{n-1}\hat{U}_l,$$

we have

$$\sum_{i=1}^{n}(\hat{U}_l - \hat{U}_l^{(i)}) = 0.$$

Therefore,

$$\sum_{i=1}^{n} \tilde{D}_{1i} = \frac{2}{(n-1)(n-2)} \sum_{i=1}^{n} \sum_{1\le l<m\le n} \sum_{k=1}^{10} DBk_{lm,i}. \tag{5.29}$$

Note that

$$|\hat{U}_l - \hat{U}_l^{(i)}| \begin{cases} \le \frac{1}{n-1}\frac{(1+h)k}{n} & if\ \hat{U}_i \ge \hat{U}_l, \\ \le \frac{1}{n-1} & if\ \hat{U}_i < \hat{U}_l, \end{cases} \tag{5.30}$$

and

$$\begin{aligned} & \textstyle\sum_{i=1}^{n} I\left(\hat{U}_l < (1+h)\frac{k}{n}\right) I\left(\hat{U}_i < \hat{U}_l\right) \\ \le & \textstyle\sum_{i=1}^{n} I\left(\hat{U}_i < (1+h)\frac{k}{n}\right) \le \frac{n+1}{n}k(1+h) = O(k). \end{aligned} \tag{5.31}$$

It follows from (5.30) and (5.31) that

$$\begin{aligned} & \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)} \textstyle\sum_{i=1}^{n} \sum_{1\le l<m\le n} DBj_{lm} \\ = & O_p(\frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{1}{n^2}khnC(\frac{k}{n},\frac{k}{n})\frac{n^2}{k^2h^2}(\frac{k^2h^2}{n^4}n+\frac{1}{n^2}k)) = O_p(\frac{h}{\sqrt{nC(\frac{k}{n},\frac{k}{n})}}\frac{k}{n}+\frac{1}{h\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1) \end{aligned}$$

for $j = 1, 2, 3, 4$, and

$$\begin{aligned} & \frac{n}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\frac{2}{n(n-1)} \textstyle\sum_{i=1}^{n} \sum_{1\le l<m\le n} DBj_{lm} \\ = & O_p(\frac{n}{nC^3(\frac{k}{n},\frac{k}{n})}\frac{1}{n^2}khnC(\frac{k}{n},\frac{k}{n})\frac{n^2}{k^2h^2}(\frac{k^2h^2}{n^4}n+\frac{1}{n^2}k)) = o_p(1) \end{aligned}$$

for $j = 5, \cdots, 10$, which imply that

$$\frac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \sum_{i=1} |\tilde{D}_{1i}| \xrightarrow{p} 0 \tag{5.32}$$

by using (5.29).

Note that $\hat{U}_i - \hat{U}_i^{(i)} = -\frac{1}{n-1}\hat{U}_i$ and $\sum_{i=1}^{n}(\hat{U}_i - \hat{U}_i^{(i)}) = -\frac{n}{n-1}\hat{U}_i$. Similar to the above derivations, we have

$$
\begin{aligned}
& \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} |DAj_{li,i}| \\
= \ & O_p(C(\tfrac{k}{n},\tfrac{k}{n})) \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \tfrac{n}{kh} \tfrac{1}{n-1} (2C_0kh)^2 \\
= \ & O_p\left(\tfrac{1}{\sqrt{nC(\frac{k}{n},\frac{k}{n})}} \tfrac{kh}{n} \right) = o_p(1)
\end{aligned}
$$

for $j = 1, 2, 3, 4$, and

$$
\begin{aligned}
& \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} |DBj_{li,i}| \\
= \ & O_p(C(\tfrac{k}{n},\tfrac{k}{n})) \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} (\tfrac{n}{kh})^2 (2C_0kh)^2 (\tfrac{1}{n-1})^2 \\
= \ & O_p(\tfrac{1}{n\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1)
\end{aligned}
$$

for $j = 5, 6, 7, 8, 9, 10$, which imply that

$$
\begin{aligned}
& \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} |DBj_{li,i}| \\
\le \ & \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} (\tfrac{n}{kh})^2 C_0^2 (2C_0kh)^2 (\tfrac{1}{n-1})^2 \\
= \ & 8C_0^4 \left(\tfrac{1}{\sqrt{nC(\frac{k}{n},\frac{k}{n})}} \right)^3 \to 0,
\end{aligned}
$$

i.e.,

$$
\frac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \sum_{i=1} |\tilde{D}_{2i}| \xrightarrow{p} 0. \tag{5.33}
$$

Further we can show that

$$
\begin{aligned}
& \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \textstyle\sum_{i=1}^n \sum_{1\le l<m\le n} |DAj_{ii,i}| \\
= \ & O_p(C(\tfrac{k}{n},\tfrac{k}{n})) \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}} \tfrac{2}{(n-1)(n-2)} \tfrac{n}{kh} (2C_0kh) \tfrac{1}{n-1} \\
= \ & O_p(\tfrac{1}{n\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1)
\end{aligned}
$$

for $j = 1, 2, 3, 4$,

$$\begin{aligned}
&\tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\tfrac{2}{(n-1)(n-2)}\textstyle\sum_{i=1}^{n}\sum_{1\le l<m\le n}|DBj_{ii,i}| \\
= \;& O_p(C(\tfrac{k}{n},\tfrac{k}{n}))\tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\tfrac{2}{(n-1)(n-2)}(\tfrac{n}{kh})^2(2C_0kh)\tfrac{1}{(n-1)^2} \\
= \;& O_P(\tfrac{1}{nkh\sqrt{nC(\frac{k}{n},\frac{k}{n})}}) = o_p(1)
\end{aligned}$$

for $j = 1, 2, 3, 4$, and

$$\begin{aligned}
&\tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\tfrac{2}{(n-1)(n-2)}\textstyle\sum_{i=1}^{n}\sum_{1\le l<m\le n}|DBj_{ii,i}| \\
\le \;& \tfrac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\tfrac{2}{(n-1)(n-2)}(\tfrac{n}{kh})^2(2C_0kh)\tfrac{1}{(n-1)^2} \\
= \;& 4C_0\tfrac{1}{kh}(\tfrac{1}{\sqrt{nC(\frac{k}{n},\frac{k}{n})}})^3 \to 0
\end{aligned}$$

for $j = 5, 6, 7, 8, 9, 10$, which imply that

$$\frac{n-1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\sum_{i=1}|\tilde{D}_{3i}| \xrightarrow{p} 0. \tag{5.34}$$

Hence, (5.25) follows from (5.32)–(5.34), i.e., the lemma holds.

□

Lemma 5.4.3. *Under conditions A1)–A3), we have*

$$\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\hat{Z}_i^2(\theta_0) \xrightarrow{P} \sigma^2 \tag{5.35}$$

as $n \to \infty$, and

$$\max_{1\le i\le n}|\hat{Z}_i(\theta_0)| = o_p\left(\sqrt{nC^3\left(\frac{k}{n},\frac{k}{n}\right)}\right). \tag{5.36}$$

Proof. According to the definition, we have

$$\begin{aligned}
\tfrac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}\hat{Z}_i^2(\theta_0) \;=\; & \tfrac{n}{nC^3(\frac{k}{n},\frac{k}{n})}\hat{T}_n^2(\theta_0) + \tfrac{2(n-1)}{nC^3(\frac{k}{n},\frac{k}{n})}\hat{T}_n(\theta_0)\textstyle\sum_{i=1}^{n}[\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0)] \\
& + \tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}[\hat{T}_n(\theta_0) - \hat{T}_n^{(i)}(\theta_0)]^2.
\end{aligned}$$

From Lemma 5.4.1 and (5.25), we conclude that

$$\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\hat{Z}_i^2(\theta_0)=\frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}[\hat{T}_n(\theta_0)-\hat{T}_n^{(i)}(\theta_0)]^2+o_p(1).$$

Therefore, to prove (5.35), we only need to show that, as $n\to\infty$

$$\frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}[\hat{T}_n(\theta_0)-\hat{T}_n^{(i)}(\theta_0)]^2\xrightarrow{p}\sigma^2. \tag{5.37}$$

By (5.26), we have

$$\begin{aligned}
&\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}[\hat{T}_n(\theta_0)-\hat{T}_n^{(i)}(\theta_0)]^2\\
=\;&\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}D_{1i}^2+\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\{D_{3i}-D_{2i}\}^2+\tfrac{2(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}D_{1i}\{D_{3i}-D_{3i}\}.
\end{aligned} \tag{5.38}$$

It follows from (5.26) and (5.28) that

$$D_{1i}=\tfrac{2}{(n-1)(n-2)}\textstyle\sum_{1\le l<m\le n}\sum_{k=1}^{4}DAk_{lm,i}+\tfrac{2}{(n-1)(n-2)}\sum_{1\le l<m\le n}\sum_{k=1}^{10}DBk_{lm,i}.$$

Thus,

$$\begin{aligned}
&\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}D_{1i}^2\\
=\;&\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}\left(\tfrac{2}{(n-1)(n-2)}\sum_{1\le l<m\le n}\sum_{k=1}^{4}DAk_{lm,i}\right)^2\\
&+\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}\left(\tfrac{2}{(n-1)(n-2)}\sum_{1\le l<m\le n}\sum_{k=1}^{10}DBk_{lm,i}\right)^2\\
&+2\tfrac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\textstyle\sum_{i=1}^{n}\left(\tfrac{2}{(n-1)(n-2)}\sum_{1\le l<m\le n}\sum_{k=1}^{4}DAk_{lm,i}\right)\times\\
&\left(\tfrac{2}{(n-1)(n-2)}\textstyle\sum_{1\le l<m\le n}\sum_{k=1}^{10}DBk_{lm,i}\right)\\
:=\;&I_1+I_2+I_3.
\end{aligned} \tag{5.39}$$

Using the fact that $|\frac{1-\frac{n}{k}\hat{U}_l}{h}|\le 1$ and $|\frac{1-\frac{n}{k}\hat{U}_m}{h}|\le 1$ uniformly for all $l,m=1,\cdots,n$, we have

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n}(\hat{U}_l - \hat{U}_l^{(i)})(\hat{U}_{l'} - \hat{U}_{l'}^{(i)}) = \frac{\hat{U}_l \wedge \hat{U}_{l'}}{(n-1)^2} - \frac{\hat{U}_l \hat{U}_{l'}}{(n-1)^2} = \frac{k}{n^3}(1 + o(\epsilon_n)), \\ \frac{1}{n}\sum_{i=1}^{n}(\hat{U}_l - \hat{U}_l^{(i)})(\hat{V}_{l'} - \hat{V}_{l'}^{(i)}) = \frac{1}{(n-1)^2}[\hat{F}(X_l, Y_{l'}) - \hat{F}_1(X_l)\hat{F}_2(Y_{l'})], \end{cases} \tag{5.40}$$

where ϵ_n's are constants and $\epsilon_n \to 0$.

First look at the first term in I_1. Using (5.40), we have

$$\begin{aligned} & \frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\{\frac{2}{n(n-1)}\sum_{1\le l<m\le n} DA1_{lm}\}^2 \\ = \quad & \frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\frac{1}{n^2(n-1)^2}\sum_{l\neq m}\sum_{l'\neq m'} B_{lm} g_h(\hat{U}_l) G_h(\hat{U}_m) G_h(\hat{V}_l) G_h(\hat{V}_m) \times \\ & B_{l'm'} g_h(\hat{U}_{l'}) G_h(\hat{U}_{m'}) G_h(\hat{V}_{l'}) G_h(\hat{V}_{m'}) \frac{n^2}{k^2h^2}\frac{k}{n^2}(1 + o_p(1)) \\ = \quad & c_0\{2\int_0^1 H(1,v) H_{12}(1,v)\, dv - (1+\theta_0) H_1(1,1)\}^2 + o_p(1). \end{aligned}$$

Similarly we can show the convergence for other terms in I_1, I_2, I_3, which lead to

$$I_1 \xrightarrow{p} \lim_{n\to\infty} E(W_{n2} + W_{n3})^2, \quad I_2 \xrightarrow{p} 0, \quad I_3 \xrightarrow{p} 0,$$

i.e.,

$$\frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n} D_{1i}^2 \xrightarrow{p} \lim_{n\to\infty} E\{W_{n2} + W_{n3}\}^2 \tag{5.41}$$

by using (5.39).

It is straightforward to verify that

$$\begin{aligned} & \frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\{\frac{2}{n(n-1)}\sum_{l=1}^{n} A_{li}^{(i)}\}^2 \\ = \quad & \frac{4}{n^3C^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{m} B_{li}B_{mi} G_h(U_l) G_h(V_l) G_h(U_i) G_h(V_i) G_h(U_m) G_h(V_m)\{1 + o_p(1)\} \\ \xrightarrow{p} \quad & 4\int\cdots\int\{sgn((u_1 - u_3)(v_1 - v_3)) - \theta_0\}\{sgn((u_2 - u_3)(v_2 - v_3)) - \theta_0\} \times \\ & dH(u_1, v_1) dH(u_2, v_2) dH(u_3, v_3) \\ = \quad & 4\int_0^1\int_0^1\{4H(u_3, v_3) - 2H(u_3, 1) - 2H(1, v_3) + 1 - \theta_0\}^2\, dH(u_3, v_3) \\ = \quad & \lim_{n\to\infty} EW_{n1}^2. \end{aligned}$$

Further we can show that

$$
\begin{aligned}
\frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^n\{D_{3i}-D_{2i}\}^2 &= \frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^n\{\frac{2}{n(n-1)}\sum_{l=1}^n A_{li}^{(i)}\}^2+o_p(1) \\
&= \lim_{n\to\infty} EW_{n1}^2+o_p(1)
\end{aligned}
\tag{5.42}
$$

and

$$
\frac{(n-1)^2}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^n D_{1i}\{D_{3i}-D_{2i}\} \xrightarrow{p} \lim_{n\to\infty} E\{W_{n1}(W_{n2}+W_{n3})\}. \tag{5.43}
$$

Hence, (5.37) follows from (5.38), (5.41)–(5.43), i.e., (5.35) follows.

Equation (5.36) can be derived similarly by examining all terms as above and we skip details.

□

Proof of Theorem 5.2.1. Denote $\bar{Z}_n(\theta_0)=\frac{1}{n}\sum_{i=1}^n\hat{Z}_i(\theta_0)$, $S_n=\frac{1}{n}\sum_{i=1}^n\hat{Z}_i^2(\theta_0)$, and $g(\lambda)=\frac{1}{n}\sum_{i=1}^n\frac{\hat{Z}_i(\theta_0)}{1+\lambda\hat{Z}_i(\theta_0)}$. Then we have

$$
\begin{aligned}
0=|g(\lambda)| &= \left|\frac{1}{n}\sum_{i=1}^n\hat{Z}_i(\theta_0)-\frac{\lambda}{n}\sum_{i=1}^n\frac{\hat{Z}_i^2(\theta_0)}{1+\lambda\hat{Z}_i(\theta_0)}\right| \\
&\ge \left|\frac{\lambda}{n}\sum_{i=1}^n\frac{\hat{Z}_i^2(\theta_0)}{1+\lambda\hat{Z}_i(\theta_0)}\right|-\left|\frac{1}{n}\sum_{i=1}^n\hat{Z}_i(\theta_0)\right| \\
&\ge \frac{|\lambda||\frac{1}{n}\sum_{i=1}^n\hat{Z}_i^2(\theta_0)|}{1+|\lambda|\max_{1\le i\le n}|\hat{Z}_i(\theta_0)|}-\left|\frac{1}{n}\sum_{i=1}^n\hat{Z}_i(\theta_0)\right|,
\end{aligned}
$$

which implies

$$
\frac{|\lambda||\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^n\hat{Z}_i^2(\theta_0)|}{1+|\lambda|\max_{1\le i\le n}|\hat{Z}_i(\theta_0)|}\le\left|\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^n\hat{Z}_i(\theta_0)\right|.
$$

From Lemma 5.4.2 and Lemma 5.4.3, we have

$$
|\lambda|=O_p\left(\frac{1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\right). \tag{5.44}
$$

Let $\gamma_i = \lambda \hat{Z}_i(\theta_0)$, then by (5.36) and (5.44), we have

$$\max_{1 \le i \le n} |\gamma_i| = o_p(1). \tag{5.45}$$

Note that

$$\begin{aligned}
0 = g(\lambda) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{Z}_i(\theta_0)}{1+\gamma_i} = \frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i(\theta_0) - \frac{\lambda}{n}\sum_{i=1}^{n} \hat{Z}_i^2(\theta_0) + \frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i(\theta_0)\frac{\gamma_i^2}{1+\gamma_i} \\
&= \bar{Z}_n(\theta_0) - \lambda S_n + \frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i(\theta_0)\frac{\gamma_i^2}{1+\gamma_i}.
\end{aligned}$$

It follows from (5.35), (5.36) and (5.45) that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i(\theta_0)\frac{\gamma_i^2}{1+\gamma_i} \le \frac{1}{n}\sum_{i=1}^{n} \hat{Z}_i^2(\theta_0)\lambda^2 \max_{1\le i\le n} \hat{Z}_i(\theta_0)\frac{1}{1+\gamma_i} = o_p\left(\frac{1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\right).$$

Therefore, $\lambda S_n = \bar{Z}_n(\theta_0) + o_p\left(\frac{1}{\sqrt{nC^3(\frac{k}{n},\frac{k}{n})}}\right)$. By Lemma 5.4.2 and Lemma 5.4.3, we have

$$\begin{aligned}
l(\theta_0) &= 2\sum_{i=1}^{n} \log(1+\lambda\hat{Z}_i(\theta_0)) \\
&= 2\sum_{i=1}^{n} \lambda\hat{Z}_i(\theta_0) - \sum_{i=1}^{n} \lambda^2\hat{Z}_i^2(\theta_0) + o_p(1) \\
&= \frac{\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\left(\sum_{i=1}^{n} \hat{Z}_i(\theta_0)\right)^2}{\frac{1}{nC^3(\frac{k}{n},\frac{k}{n})}\sum_{i=1}^{n} \hat{Z}_i^2(\theta_0)} + o_p(1) \to \chi_1^2.
\end{aligned}$$

$\square$

Table 5.1: Coverage probabilities are computed for the bootstrap method based on the smoothing estimator $\tilde{\theta}$ (i.e., solving $\hat{T}_n(\theta) = 0$) and the empirical likelihood method for $k = 50, 100, 150$, and $h = \delta\{\sum_{i=1}^n I(\hat{U}_i \le \frac{k}{n}, \hat{V}_i \le \frac{k}{n})\}^{-1/3}$ with $\delta = 0.5, 1, 1.5$. We take $\rho = 0.5$ for both normal copula and elliptical distribution.

(k, δ)	Distribution	α	$I^B_{0.90}$	$I^B_{0.95}$	$I^{EL}_{0.90}$	$I^{LE}_{0.95}$
$(50, 0.5)$	Normal Copula		0.853	0.911	0.889	0.954
$(50, 1.0)$	Normal Copula		0.869	0.920	0.907	0.970
$(50, 1.5)$	Normal Copula		0.854	0.917	0.903	0.964
$(50, 0.5)$	Elliptical	1	0.887	0.921	0.894	0.931
$(50, 1.0)$	Elliptical	1	0.874	0.925	0.893	0.939
$(50, 1.5)$	Elliptical	1	0.885	0.924	0.898	0.939
$(50, 0.5)$	Elliptical	5	0.822	0.880	0.893	0.953
$(50, 1.0)$	Elliptical	5	0.829	0.892	0.905	0.959
$(50, 1.5)$	Elliptical	5	0.832	0.888	0.898	0.956
$(100, 0.5)$	Normal Copula		0.823	0.881	0.815	0.889
$(100, 1.0)$	Normal Copula		0.824	0.882	0.809	0.897
$(100, 1.5)$	Normal Copula		0.816	0.884	0.801	0.883
$(100, 0.5)$	Elliptical	1	0.882	0.939	0.889	0.938
$(100, 1.0)$	Elliptical	1	0.879	0.936	0.889	0.944
$(100, 1.5)$	Elliptical	1	0.872	0.921	0.883	0.941
$(100, 0.5)$	Elliptical	5	0.874	0.923	0.906	0.942
$(100, 1.0)$	Elliptical	5	0.872	0.926	0.911	0.953
$(100, 1.5)$	Elliptical	5	0.873	0.925	0.911	0.952
$(150, 0.5)$	Normal Copula		0.742	0.819	0.720	0.823
$(150, 1.0)$	Normal Copula		0.721	0.798	0.701	0.800
$(150, 1.5)$	Normal Copula		0.693	0.776	0.673	0.777
$(150, 0.5)$	Elliptical	1	0.888	0.933	0.895	0.944
$(150, 1.0)$	Elliptical	1	0.890	0.937	0.898	0.947
$(150, 1.5)$	Elliptical	1	0.882	0.936	0.888	0.944
$(150, 0.5)$	Elliptical	5	0.885	0.942	0.919	0.954
$(150, 1.0)$	Elliptical	5	0.887	0.937	0.919	0.962
$(150, 1.5)$	Elliptical	5	0.889	0.934	0.920	0.954

Table 5.2: Confidence intervals with level 0.9 and 0.95 are computed for the bootstrap method based on the smoothing estimator $\tilde{\theta}$ (i.e., solving $\hat{T}_n(\theta) = 0$) and the empirical likelihood method for $k = 60, 70, 80, 90, 100$, and $h = \delta\{\sum_{i=1}^n I(\hat{U}_i \le \frac{k}{n}, \hat{V}_i \le \frac{k}{n})\}^{-1/3}$ with $\delta = 0.5, 1, 1.5$.

(k, δ)	$\tilde{\theta}$	$I_{0.90}^B$	$I_{0.95}^B$	$I_{0.90}^{EL}$	$I_{0.95}^{LE}$
$(60, 0.5)$	0.005	(-0.332, 0.263)	(-0.419, 0.303)	(-0.35, 0.37)	(-0.45, 0.48)
$(60, 1.0)$	0.013	(-0.313, 0.245)	(-0.379, 0.290)	(-0.37, 0.30)	(-0.32, 0.37)
$(60, 1.5)$	0.036	(-0.275, 0.241)	(-0.319, 0.304)	(-0.23, 0.30)	(-0.29, 0.37)
$(70, 0.5)$	-0.012	(-0.335, 0.179)	(-0.383, 0.236)	(-0.26, 0.21)	(-0.31, 0.26)
$(70, 1.0)$	0.014	(-0.276, 0.199)	(-0.327, 0.237)	(-0.23, 0.24)	(-0.28, 0.30)
$(70, 1.5)$	0.040	(-0.232, 0.253)	(-0.261, 0.297)	(-0.20, 0.27)	(-0.25, 0.33)
$(80, 0.5)$	0.057	(-0.218, 0.265)	(-0.254, 0.315)	(-0.24, 0.31)	(-0.30, 0.37)
$(80, 1.0)$	0.070	(-0.169, 0.254)	(-0.212, 0.310)	(-0.16, 0.29)	(-0.21, 0.34)
$(80, 1.5)$	0.79	(-0.147, 0.262)	(-0.179, 0.298)	(-0.12, 0.29)	(-0.16, 0.34)
$(90, 0.5)$	0.148	(-0.044, 0.372)	(-0.089, 0.404)	(0.00, 0.40)	(-0.05, 0.44)
$(90, 1.0)$	0.120	(-0.101, 0.323)	(-0.131, 0.362)	(-0.05, 0.33)	(-0.09, 0.37)
$(90, 1.5)$	0.104	(-0.097, 0.288)	(-0.124, 0.316)	(-0.07, 0.31)	(-0.10, 0.35)
$(100, 0.5)$	0.138	(-0.046, 0.320)	(-0.089, 0.363)	(-0.03, 0.31)	(-0.07, 0.34)
$(100, 1.0)$	0.127	(-0.058, 0.312)	(-0.081, 0.354)	(-0.02, 0.32)	(-0.06, 0.36)
$(100, 1.5)$	0.106	(-0.064, 0.266)	(-0.092, 0.306)	(-0.05, 0.30)	(-0.08, 0.33)

Appendices

APPENDIX A
THE MATRIX IN LEMMA 3.6.14

The matrix V in Lemma 3.6.14 is given by

$$\begin{bmatrix} E[w_{x,t}U_{x,t}(\boldsymbol{\psi}_x^0)U_{x,t}^T(\boldsymbol{\psi}_x^0)] & 0 & 0 & 0 & 0 & 0 \\ E(U_{x,t})J_1^T & 0 & f_{\eta_{x,t}}(\theta_1^0) & 0 & 0 & 0 \\ E(U_{x,t})J_3^T & 0 & -\theta_1^0 f_{\eta_{x,t}}(\theta_1^0) & 0 & -1 & 0 \\ 0 & E[w_{y,t}U_{y,t}(\boldsymbol{\psi}_y^0)U_{y,t}^T(\boldsymbol{\psi}_y^0)] & 0 & 0 & 0 & 0 \\ 0 & E(U_{y,t})J_2^T & 0 & f_{\eta_{y,t}}(\theta_2^0) & 0 & 0 \\ 0 & E(U_{y,t})J_4^T & 0 & -\theta_2^0 f_{\eta_{y,t}}(\theta_2^0) & 0 & -1 \\ F_x'(\theta_1^0,\theta_2^0)E(U_{x,t})J_1^T & F_y'(\theta_1^0,\theta_2^0)E(U_{y,t})J_2^T & F_x'(\theta_1^0,\theta_2^0) & F_y'(\theta_1^0,\theta_2^0) & \alpha\frac{\theta_5^0\theta_4^0}{(\theta_3^0)^2} & -\alpha\frac{\theta_5^0}{\theta_3^0} \end{bmatrix}$$

where $U_{x,t} = \left[\frac{1}{\sqrt{h_{x,t}}}\frac{\partial \varepsilon_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}, \frac{1}{\sqrt{2h_{x,t}}}\frac{\partial h_{x,t}(\boldsymbol{\psi}_x^0)}{\partial \boldsymbol{\psi}_x}\right], U_{y,t} = \left[\frac{1}{\sqrt{h_{y,t}}}\frac{\partial \varepsilon_{y,t}(\boldsymbol{\psi}_y^0)}{\partial \boldsymbol{\psi}_y}, \frac{1}{\sqrt{2h_{y,t}}}\frac{\partial h_{y,t}(\boldsymbol{\psi}_y^0)}{\partial \boldsymbol{\psi}_y}\right]$,
and

$$J_1 = -f_{\eta_{x,t}}(\theta_1^0)[2\theta_1^0, -\sqrt{2}(\theta_1^0)^2], \quad J_2 = -f_{\eta_{y,t}}(\theta_2^0)[2\theta_2^0, -\sqrt{2}(\theta_2^0)^2],$$
$$J_3 = [2\theta_3^0, \sqrt{2}E(\eta_{x,t}^2\mathbb{1}[\eta_{x,t} \le \theta_1^0])] - \theta_1^0 J_1, \quad J_4 = [2\theta_3^0, \sqrt{2}E(\eta_{y,t}^2\mathbb{1}[\eta_{y,t} \le \theta_1^0])] - \theta_1^0 J_2.$$

APPENDIX B

AN APPLICATION OF RELATIVE RISK MEASURE TO SYSTEMIC RISK

In this section, we will apply the relative risk measure to the network-based model of a financial system. The purpose is to illustrate how pairwise risk measure is used to construct network-based risk measures for systemic risk.

A financial system can be modeled on a directed graph with vertices representing individual institutes and an edge between two vertices indicating directed risk flows. Moreover, observed data are generated on each vertex across time. More precisely, let $G = (V, E)$ be a directed graph where $V = \{1, 2, \ldots, N\} \subset \mathbb{N}$ is the set of vertices and $E = \{(i, j)|i, j \in V\}$ is the set of directed edges. To facilitate statistical modeling, the edges are supposed to have weights and we use adjacency matrix to record the weights. That is, we use $A = (\theta_{ij})_{i,j \in V} \in \mathcal{M}_N(\mathbb{R})$, where θ_{ij} is the weight on the edge from i to j, and thus A is a matrix of parameters in statistical modeling.

As an example, we choose 33 financial institutes as the vertices which are listed in the Table B.1 and use relative risk measure as the risk flows. Let θ_{ij} be the relative risk measure to institute i from institute j (j is the benchmark). We let $\theta_{ij} = 0$ when $i = j$ since relative risk makes no sense for an institute to itself. Furthermore, suppose $\{\mathbf{y}_t = (y_{kt}; k \in V)^T\}_{t \le n}$ is the sample data generated on the vertex set V. The time series of daily stock loss data of the 33 financial institutes from December 2001 to May 2008 (see Table B.1). We predict the one-day-ahead relative risk measures of the financial system.

For systemic risk measures, one intuitive and standard construction is based on degree distributions of the graph models. There are usually three levels of measures: 1) between two vertices, 2) between a vertex and the system, and 3) global.

- **Pairwise Risk Influence:** The edges measure the proportion of extreme co-movements occurring between two vertices, which serves as pairwise measures.

$$\theta_{ij} = e_i E e_j^T.$$

- **Local-to-System and System-to-Local Risk Influence:** We use $+$ to indicate the inflows and $-$ the outflows of a network at a vertex. The risk influence from/to system to/from vertex i are defined as

$$\Theta_i^+ = \frac{1}{N} \sum_{j=1, j\neq i}^{N} \theta_{ij}, \quad \text{and} \quad \Theta_i^- = \frac{1}{N} \sum_{j=1, j\neq i}^{N} \theta_{ji}.$$

- **Global Risk Influence:** Global risk influence is a measure excluding the self risk influence on each vertex. We define it as

$$\Theta = \frac{1}{N} \sum_{i,j=1, i\neq j}^{N} \theta_{ji}.$$

Based on the above modeling, we estimate the adjacency matrix by a two-step approach: First, we fit a AR(1)-GARCH(1,1) model to stock returns of each institute. The reason why we choose AR(1)-GARCH(1,1) model is similar to that in Section 3.5. Second, we use the estimated model and the residuals to predict the one-day-ahead (conditional) relative risk in (3.9) for each pair of two institutes. The estimated adjacency matrix consists of all estimated θ_{ij}s as (i, j)-th elements, with diagonal entries being zeros. Note that we set risk level 0.05 in the relative risk measure for all vertices.

To visualize the overall estimation, we display the adjacency matrix in Figure B.1. The cells are of form (institute i, institute j) which are θ_{ij}'s of the directed edges to i from j. However, instead of displaying the estimated values, we use color to describe the extent of the risk measures for a good comparison. The color density indicates the value of relative risk measure: darker red means lower value while lighter yellow means greater value.

Figure B.1 shows that most cells have darker red which means lower risk measures, but a few cells, say, cells (SAF, ALL),(SAF, CB),(LEH, ALL) and (LEH,USB), have lighter yellow which means high risk measures. For instance, (SAF,ALL), i.e. the relative risk measure to SAF from ALL, has the greatest value among all the pairwise risk flows since it has the lightest color.

To understand the Local-to-System (Out-Risk) and System-to-Local (In-Risk) risk measures, we scale the adjacency matrix by column or row. Figure B.2 shows the adjacency matrix by row scaling, that is, we divide each cell in the matrix by the sum of the row where the cell is located. Therefore, the sum of each row is 1, and the value of each cell (i, j) is between 0 and 1, which is the percent of institute j contributing to Θ_i^+. We observe that ALL and USB have the highest percentages among most Θ_i^+, which implies that these two companies are essential to the System-to-Local risk measures within the network. The estimated values are summarized in Table B.1.

To display the Local-to-System risk measures Θ_i^+ for each companies, we use both the size and color of the vertices to show the extent of Θ_i^+ in Figure B.4: larger size and lighter yellow means greater value. The directed edges of the network in Figure B.4 correspond to elements of the estimated adjacency matrix in Figure B.1; however, we only sketch those edges whose weights are above the median of all pairwise risk measures with both more width and black meaning greater value. We observe that LET and SAF have the greatest In-Risk Θ_i^+ since they have the largest size and lightest color.

Similarly, we can perform the same analysis for the Local-to-System Θ_i^- ,and results are presented in Figures B.3 and B.5. We observe that SAF and LEH have the highest percentages among most Θ_i^i in Figure B.3, which implies that these two companies encounter the highest percentages in System-to-Local risk measures within the network. In addition, ALL and USB have the greatest Out-Risk Θ_i^- since they have the largest size and lightest color in Figure B.5. Note that the results exhibited in Figure B.2 are consistent with the ones in Figure B.5; the results exhibited in Figure B.3 are consistent with the ones in Figure

B.4.

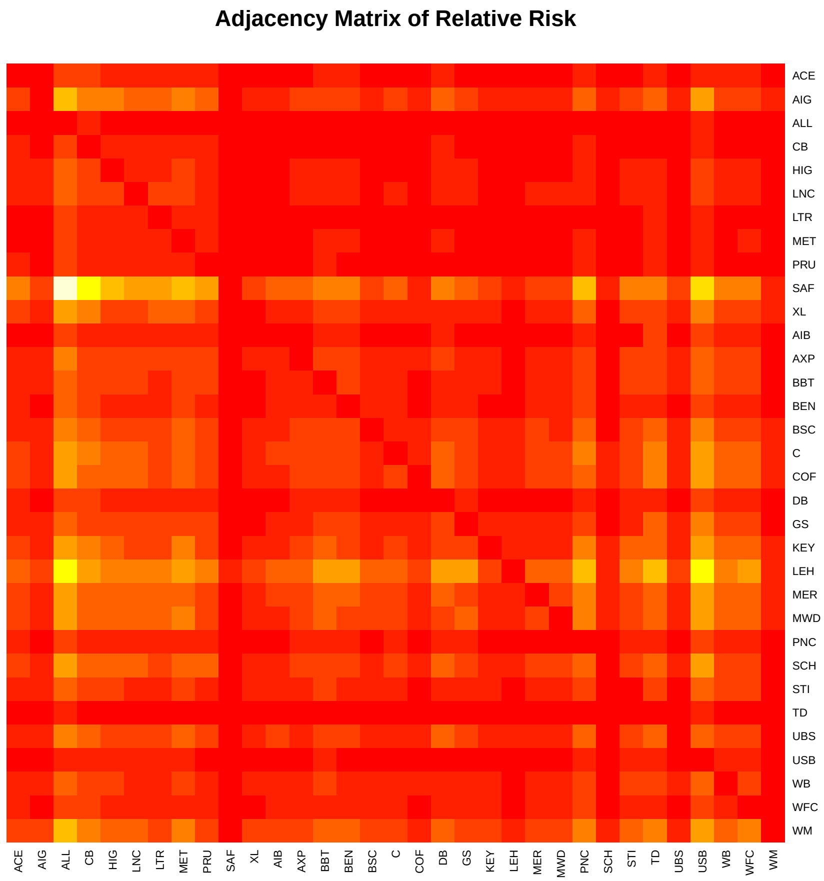


Figure B.1: Estimated Adjacency Matrix.

Firms	Ticker	In-Risk Θ_i^+	Out-Risk Θ_i^-
ACE Ltd	ACE	0.2683251	0.4500215
AIG	AIG	0.6920917	0.3335920
Allstate Corporation	ALL	0.1364261	1.0959577
Chubb Corporation	CB	0.2182572	0.7632567
Hartford Financial Services	HIG	0.3253704	0.6369664
Lincoln National	LNC	0.3738475	0.5894946
Loews Corporation	LTR	0.2112635	0.5612856
MetLife	MET	0.2488326	0.7353998
Prudential Financial	PRU	0.2322405	0.5416503
SAFECO Corporation	SAF	1.1309755	0.1447151
XL Capital	XL	0.5944673	0.2707453
Allied Irish Banks	AIB	0.2947133	0.3667621
American Express	AXP	0.4926042	0.4320578
BB&T Corporation	BBT	0.4550936	0.5646702
Franklin Resources	BEN	0.3645421	0.5277011
Bear Stearns	BSC	0.5864601	0.3369254
Citigroup	C	0.7105274	0.3850313
Capital One	COF	0.6901228	0.2497780
Deutsche Bank	DB	0.2911164	0.5639638
Goldman Sachs	GS	0.5067724	0.4503810
Keycorp	KEY	0.7055933	0.3110573
Lehman Brothers	LEH	1.1009892	0.2226547
Merrill Lynch	MER	0.7116743	0.3773102
Morgan Stanley	MWD	0.7024409	0.3577850
PNC Financial	PNC	0.3054912	0.7057088
Charles Schwab	SCH	0.6637684	0.2047582
SunTrust Banks	STI	0.4205891	0.5208059
Toronto Dominion Bank	TD	0.1467565	0.6922985
UBS	UBS	0.5943143	0.2927379
Us Bancorp	USB	0.2166317	0.9698869
Wachovia	WB	0.4674905	0.5444146
Wells Fargo	WFC	0.3532215	0.5865574
Washington Mutual	WM	0.7763328	0.2030125

Table B.1: 33 Financial Institutes with Global Risk Measure 15.98934.

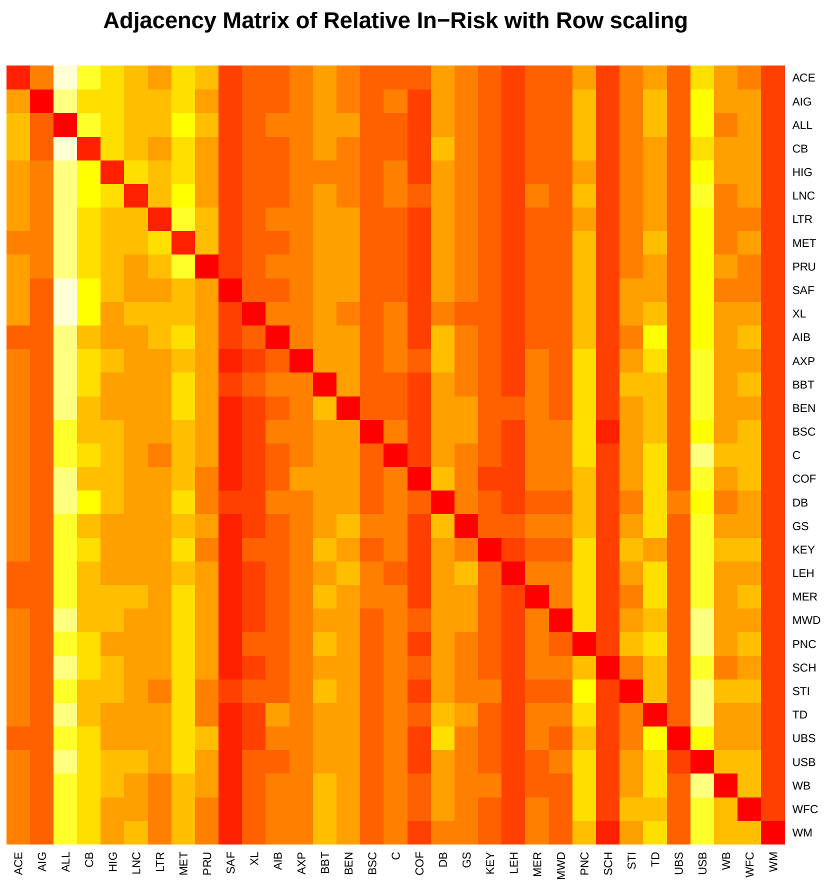


Figure B.2: Adjacency Matrix with Row Scaling for In-Risk.

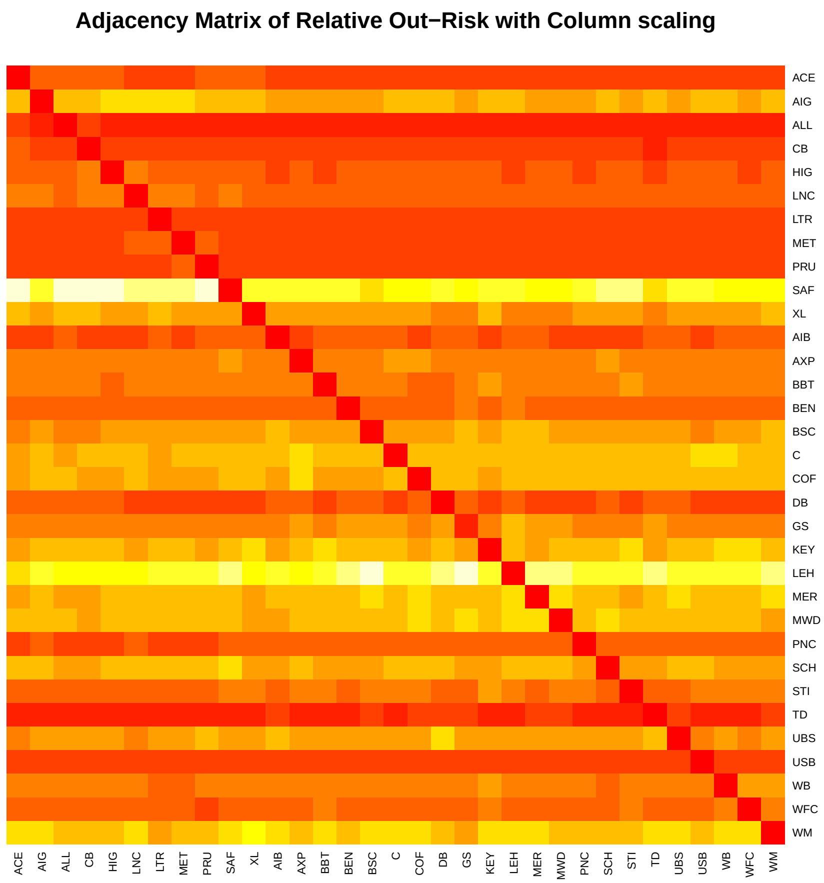


Figure B.3: Adjacency Matrix with Column Scaling for Out-Risk.

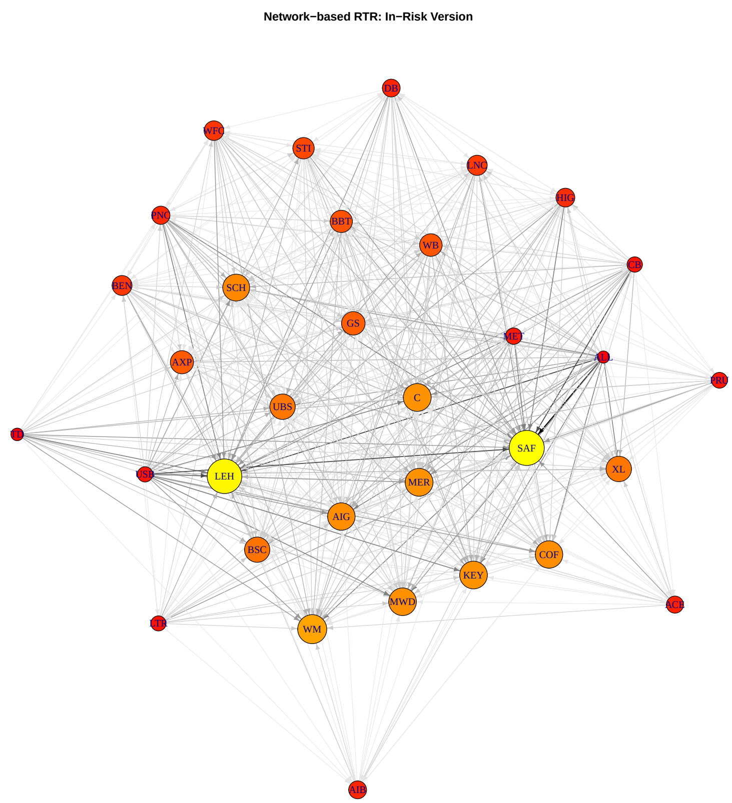


Figure B.4: Network-based In-Risk.

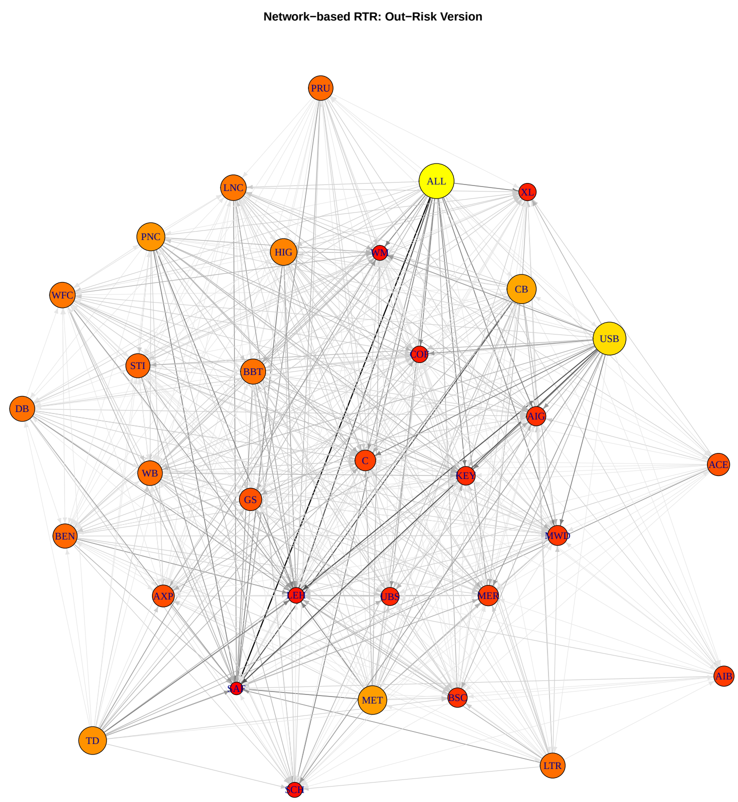


Figure B.5: Network-based Out-Risk.

REFERENCES

[1] A. B. Owen, *Empirical likelihood*. CRC press, 2001.

[2] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *The Annals of Statistics*, pp. 300–325, 1994.

[3] M. Sklar, *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[4] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[5] H. Joe, *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[6] C. Genest, K. Ghoudi, and L.-P. Rivest, "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, pp. 543–552, 1995.

[7] F. Durante and C. Sempi, "Copula theory: An introduction," in *Copula theory and its applications*, Springer, 2010, pp. 3–31.

[8] J.-D. Fermanian, D. Radulovic, M. Wegkamp, *et al.*, "Weak convergence of empirical copula processes," *Bernoulli*, vol. 10, no. 5, pp. 847–860, 2004.

[9] X. Huang, *Statistics of bivariate extreme values*. Thesis Publishers, 1992.

[10] L. De Haan and A. Ferreira, *Extreme value theory: An introduction*. Springer Science & Business Media, 2007.

[11] M. Sibuya, "Bivariate extreme statistics, i," *Annals of the Institute of Statistical Mathematics*, vol. 11, no. 2, pp. 195–210, 1960.

[12] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Variance reduction techniques for estimating value-at-risk," *Management Science*, vol. 46, no. 10, pp. 1349–1364, 2000.

[13] S. X. Chen and P. Hall, "Smoothed empirical likelihood confidence intervals for quantiles," *The Annals of Statistics*, pp. 1166–1181, 1993.

[14] S. X. Chen, “Nonparametric estimation of expected shortfall,” *Journal of financial econometrics*, vol. 6, no. 1, pp. 87–107, 2008.

[15] R. E. Baysal and J. Staum, “Empirical likelihood for value-at-risk and expected shortfall,” *The journal of risk*, vol. 11, no. 1, p. 3, 2008.

[16] S. Y. Chun, A. Shapiro, and S. Uryasev, “Conditional value-at-risk and average value-at-risk: Estimation and asymptotics,” *Operations Research*, vol. 60, no. 4, pp. 739–756, 2012.

[17] M. E. Yaari, “The dual theory of choice under risk,” *Econometrica: Journal of the Econometric Society*, pp. 95–115, 1987.

[18] P. Artzner, “Application of coherent risk measures to capital requirements in insurance,” *North American Actuarial Journal*, vol. 3, no. 2, pp. 11–25, 1999.

[19] C. Acerbi, “Spectral measures of risk: A coherent representation of subjective risk aversion,” *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1505–1518, 2002.

[20] D. Tasche, “Expected shortfall and beyond,” *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1519–1533, 2002.

[21] B. L. Jones and R. Zitikis, “Empirical estimation of risk measures and related quantities,” *North American Actuarial Journal*, vol. 7, no. 4, pp. 44–54, 2003.

[22] S. S. Wang, V. R. Young, and H. H. Panjer, “Axiomatic characterization of insurance prices,” *Insurance: Mathematics and economics*, vol. 21, no. 2, pp. 173–183, 1997.

[23] S. Wang, “Insurance pricing and increased limits ratemaking by proportional hazards transforms,” *Insurance: Mathematics and Economics*, vol. 17, no. 1, pp. 43–54, 1995.

[24] ——, “Premium calculation by transforming the layer premium density,” *ASTIN Bulletin: The Journal of the IAA*, vol. 26, no. 1, pp. 71–92, 1996.

[25] ——, “An actuarial index of the right-tail risk,” *North American Actuarial Journal*, vol. 2, no. 2, pp. 88–101, 1998.

[26] J. L. Wirch and M. R. Hardy, “A synthesis of risk measures for capital adequacy,” *Insurance: mathematics and Economics*, vol. 25, no. 3, pp. 337–347, 1999.

[27] A. Necir and D. Meraghni, “Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts,” *Insurance: Mathematics and economics*, vol. 45, no. 1, pp. 49–58, 2009.

[28] J. Haezendonck and M. Goovaerts, “A new premium calculation principle based on orlicz norms,” *Insurance: Mathematics and Economics*, vol. 1, no. 1, pp. 41–53, 1982.

[29] M. Goovaerts, D. Linders, K. Van Weert, and F. Tank, “On the interplay between distortion, mean value and haezendonck–goovaerts risk measures,” *Insurance: Mathematics and Economics*, vol. 51, no. 1, pp. 10–18, 2012.

[30] D. Duffie and J. Pan, “An overview of value at risk,” *The Journal of derivatives*, vol. 4, no. 3, pp. 7–49, 1997.

[31] ——, “Analytical value-at-risk with jumps and credit risk,” *Finance and Stochastics*, vol. 5, no. 2, pp. 155–180, 2001.

[32] J. Philippe, “Value at risk: The new benchmark for managing financial risk,” *NY: McGraw-Hill Professional*, 2001.

[33] S. Kou, X. Peng, and C. C. Heyde, “External risk measures and basel accords,” *Mathematics of Operations Research*, vol. 38, no. 3, pp. 393–417, 2013.

[34] S. Kou and X. Peng, “On the measurement of economic tail risk,” *Operations Research*, vol. 64, no. 5, pp. 1056–1072, 2016.

[35] P. Glasserman, P. Heidelberger, and P. Shahabuddin, “Variance reduction techniques for estimating value-at-risk,” *Management Science*, vol. 46, no. 10, pp. 1349–1364, 2000.

[36] J. Berkowitz, P. Christoffersen, and D. Pelletier, “Evaluating value-at-risk models with desk-level data,” *Management Science*, vol. 57, no. 12, pp. 2213–2227, 2011.

[37] K. Kuester, S. Mittnik, and M. S. Paolella, “Value-at-risk prediction: A comparison of alternative strategies,” *Journal of Financial Econometrics*, vol. 4, no. 1, pp. 53–89, 2006.

[38] S. X. Chen and P. Hall, “Smoothed empirical likelihood confidence intervals for quantiles,” *The Annals of Statistics*, pp. 1166–1181, 1993.

[39] R. E. Baysal and J. Staum, “Empirical likelihood for value-at-risk and expected shortfall,” *The journal of risk*, vol. 11, no. 1, p. 3, 2008.

[40] K.-L. Xu, “Model-free inference for tail risk measures,” *Econometric Theory*, vol. 32, no. 01, pp. 122–153, 2016.

[41] R. F. Engle and S. Manganelli, "Caviar: Conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, vol. 22, no. 4, pp. 367–381, 2004.

[42] P. Hall and C. C. Heyde, *Martingale limit theory and its application*. Academic press, 1980.

[43] A. Owen, "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, pp. 90–120, 1990.

[44] R. P. Sherman, "The limiting distribution of the maximum rank correlation estimator," *Econometrica: Journal of the Econometric Society*, pp. 123–137, 1993.

[45] D. Bisias, M. Flood, A. W. Lo, and S. Valavanis, "A survey of systemic risk analytics," *Annu. Rev. Financ. Econ.*, vol. 4, no. 1, pp. 255–296, 2012.

[46] H. Chen, J. D. Cummins, K. S. Viswanathan, and M. A. Weiss, "Systemic risk and the interconnectedness between banks and insurers: An econometric analysis," *Journal of Risk and Insurance*, vol. 81, no. 3, pp. 623–652, 2014.

[47] H. Chen, J. D. Cummins, T. Sun, and M. A. Weiss, "Tail risk spillover and its contribution to systemic risk: A network analysis for global reinsurers," 2015.

[48] E. Berdin and M. Sottocornola, "Insurance activities and systemic risk," 2015.

[49] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, no. 3, pp. 535–559, 2012.

[50] G. N. Weiß and J. Mühlnickel, "Why do some insurers become systemically relevant?" *Journal of Financial Stability*, vol. 13, pp. 95–117, 2014.

[51] C. Bierth, F. Irresberger, and G. N. Weiß, "Systemic risk of insurers around the globe," *Journal of Banking & Finance*, vol. 55, pp. 232–245, 2015.

[52] V. Agarwal, S. Ruenzi, and F. Weigert, "Tail risk in hedge funds: A unique view from portfolio holdings," 2016.

[53] A. W. Ledford and J. A. Tawn, "Statistics for near independence in multivariate extreme values," *Biometrika*, vol. 83, no. 1, pp. 169–187, 1996.

[54] ——, "Modelling dependence within joint tail regions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 2, pp. 475–499, 1997.

[55] L. Peng, Y. Qi, R. Wang, and J. Yang, "Jackknife empirical likelihood method for some risk measures and related quantities," *Insurance: Mathematics and Economics*, vol. 51, no. 1, pp. 142–150, 2012.

[56] L. Peng, X. Wang, and Y. Zheng, "Empirical likelihood inference for haezendonck-goovaerts risk measure," *European Actuarial Journal*, vol. 5, no. 2, pp. 427–445, 2015.

[57] X. Wang and L. Peng, "Inference for intermediate haezendonck–goovaerts risk measure," *Insurance: Mathematics and Economics*, vol. 68, pp. 231–240, 2016.

[58] L. Peng and Y. Qi, "Smoothed jackknife empirical likelihood method for tail copulas," *Test*, vol. 19, no. 3, pp. 514–536, 2010.

[59] L. Peng, Y. Qi, and I. Van Keilegom, "Jackknife empirical likelihood method for copulas," *Test*, vol. 21, no. 1, pp. 74–92, 2012.

[60] B.-Y. Jing, J. Yuan, and W. Zhou, "Jackknife empirical likelihood," *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 1224–1232, 2009.

[61] B. Kelly and H. Jiang, "Tail risk and asset prices," *Review of Financial Studies*, vol. 27, no. 10, pp. 2841–2871, 2014.

[62] P. Gaenssler *et al.*, *Seminar on empirical processes*. Birkhäuser, 2013, vol. 9.

[63] J.-J. Cai, J. H. Einmahl, L. Haan, and C. Zhou, "Estimation of the marginal expected shortfall: The mean when a related variable is extreme," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 77, no. 2, pp. 417–442, 2015.

[64] J. H. Einmahl, L. De Haan, D. Li, *et al.*, "Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition," *The Annals of Statistics*, vol. 34, no. 4, pp. 1987–2014, 2006.

[65] H. Drees and X. Huang, "Best attainable rates of convergence for estimators of the stable tail dependence function," *Journal of Multivariate Analysis*, vol. 64, no. 1, pp. 25–46, 1998.

[66] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels, *Statistics of extremes: Theory and applications*. John Wiley & Sons, 2006.

[67] A. Owen, "Empirical likelihood ratio confidence intervals for a single functional. biometrika 75 237–249," *Mathematical Reviews (MathSciNet): MR946049 Zentralblatt MATH*, vol. 641, 1988.

[68] Y. Gong, L. Peng, and Y. Qi, "Smoothed jackknife empirical likelihood method for roc curve," *Journal of Multivariate Analysis*, vol. 101, no. 6, pp. 1520–1531, 2010.

[69] C. Francq, J.-M. Zakoïan, *et al.*, "Maximum likelihood estimation of pure garch and arma-garch processes," *Bernoulli*, vol. 10, no. 4, pp. 605–637, 2004.

[70] S. Ling, "Self-weighted and local quasi-maximum likelihood estimators for arma-garch/igarch models," *Journal of Econometrics*, vol. 140, no. 2, pp. 849–873, 2007.

[71] K. Zhu, S. Ling, *et al.*, "Global self-weighted and local quasi-maximum exponential likelihood estimators for arma–garch/igarch models," *The Annals of Statistics*, vol. 39, no. 4, pp. 2131–2163, 2011.

[72] X. Luo and P. V. Shevchenko, "The t copula with multiple parameters of degrees of freedom: Bivariate characteristics and application to risk management," *Quantitative Finance*, vol. 10, no. 9, pp. 1039–1054, 2010.

[73] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[74] V. V. Acharya, L. H. Pedersen, T. Philippon, and M. Richardson, "Measuring systemic risk," *Review of Financial Studies*, vol. 30, no. 1, pp. 2–47, 2017.

[75] C. Genest, J. Segers, *et al.*, "Rank-based inference for bivariate extreme-value copulas," *The Annals of Statistics*, vol. 37, no. 5B, pp. 2990–3022, 2009.

[76] D. Pollard, *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

[77] W. Vervaat, "Functional central limit theorems for processes with positive drift and their inverses," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 23, no. 4, pp. 245–253, 1972.

[78] H. Potter, "The mean values of certain dirichlet series, ii," *Proceedings of the London Mathematical Society*, vol. 2, no. 1, pp. 1–19, 1942.

[79] S. H. Sung, "Weak law of large numbers for arrays," *Statistics & probability letters*, vol. 38, no. 2, pp. 101–105, 1998.

[80] B. Committee *et al.*, "Basel iii: A global regulatory framework for more resilient banks and banking systems," *Basel Committee on Banking Supervision, Basel*, 2010.

[81] E. Directive, "Directive 2009/147/ec of the european parliament and of the council of 30 november 2009 on the conservation of wild birds (codified version)," *Official Journal L*, vol. 20, pp. 7–25, 2009.

[82] S. S. Test, *Finma sst technisches dokument*, 2006.

[83] J. R. Aragonés, C. Blanco, and K. Dowd, "Incorporating stress tests into market risk modeling (digest summary)," *Derivatives Quarterly*, vol. 7, no. 3, pp. 44–49, 2001.

[84] F. M. Longin, "From value at risk to stress testing: The extreme value approach," *Journal of Banking & Finance*, vol. 24, no. 7, pp. 1097–1130, 2000.

[85] L. Peng, "Estimation of the coefficient of tail dependence in bivariate extremes," *Statistics & Probability Letters*, vol. 43, no. 4, pp. 399–409, 1999.

[86] G. Draisma, H. Dress, A. Ferreira, and L. De Haan, "Bivariate tail estimation: Dependence in asymptotic independence," *Bernoulli*, pp. 251–280, 2004.

[87] Y. Goegebeur and A. Guillou, "Asymptotically unbiased estimation of the coefficient of tail dependence," *Scandinavian Journal of Statistics*, vol. 40, no. 1, pp. 174–189, 2013.

[88] E. W. Frees and E. A. Valdez, "Understanding relationships using copulas," *North American actuarial journal*, vol. 2, no. 1, pp. 1–25, 1998.

[89] W. Breymann, A. Dias, and P. Embrechts, "Dependence structures for multivariate high-frequency data in finance," 2003.

[90] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, 2015.

[91] G. G. Venter, "Tails of copulas," in *Proceedings of the Casualty Actuarial Society*, vol. 89, 2002, pp. 68–113.

[92] I. Gijbels, N. Veraverbeke, and M. Omelka, "Conditional copulas, association measures and their applications," *Computational Statistics & Data Analysis*, vol. 55, no. 5, pp. 1919–1932, 2011.

[93] B. Basrak, R. A. Davis, and T. Mikosch, "Regular variation of garch processes," *Stochastic processes and their applications*, vol. 99, no. 1, pp. 95–115, 2002.

[94] L. Hua and H. Joe, "Tail order and intermediate tail dependence of multivariate copulas," *Journal of Multivariate Analysis*, vol. 102, no. 10, pp. 1454–1471, 2011.

[95] ——, “Intermediate tail dependence: A review and some new results,” in *Stochastic Orders in Reliability and Risk*, Springer, 2013, pp. 291–311.

[96] T. Mikosch and O. Wintenberger, “The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate markov chains,” *Probability Theory and Related Fields*, vol. 159, no. 1-2, pp. 157–196, 2014.

[97] L. De Haan and S. I. Resnick, “Derivatives of regularly varying functions in rd and domains of attraction of stable distributions,” *Stochastic Processes and their Applications*, vol. 8, no. 3, pp. 349–355, 1979.

[98] S. I. Resnick, *Extreme values, regular variation and point processes*. Springer, 2013.

[99] L. Peng and Y. Qi, “Bootstrap approximation of tail dependence function,” *Journal of Multivariate Analysis*, vol. 99, no. 8, pp. 1807–1824, 2008.

[100] S. Cheng and L. Peng, “Confidence intervals for the tail index,” *Bernoulli*, pp. 751–760, 2001.

[101] C. Klüppelberg, G. Kuhn, and L. Peng, “Semi-parametric models for the multivariate tail dependence function–the asymptotically dependent case,” *Scandinavian Journal of Statistics*, vol. 35, no. 4, pp. 701–718, 2008.

[102] A. Juri and M. V. Wüthrich, “Tail dependence from a distributional point of view,” *Extremes*, vol. 6, no. 3, pp. 213–246, 2003.

[103] A. V. Asimit and B. L. Jones, “Extreme behavior of bivariate elliptical distributions,” *Insurance: Mathematics and Economics*, vol. 41, no. 1, pp. 53–61, 2007.

[104] A. Asimit, R. Gerrard, Y. Hou, and L. Peng, “Tail dependence measure for modeling financial extreme co-movements,” *J. Econometrics*, 2015.

[105] L. Allen, T. G. Bali, and Y. Tang, “Does systemic risk in the financial sector predict future economic downturns?” *Review of Financial Studies*, vol. 25, no. 10, pp. 3000–3036, 2012.

[106] D. Bisias, M. Flood, A. W. Lo, and S. Valavanis, “A survey of systemic risk analytics,” *Annu. Rev. Financ. Econ.*, vol. 4, no. 1, pp. 255–296, 2012.

[107] S. Haug, C. Klüppelberg, and L. Peng, “Statistical models and methods for dependence in insurance data,” *Journal of the Korean Statistical Society*, vol. 40, no. 2, pp. 125–139, 2011.

[108] C. Dutang, Y. Goegebeur, and A. Guillou, “Robust and bias-corrected estimation of the coefficient of tail dependence,” *Insurance: Mathematics and Economics*, vol. 57, pp. 46–57, 2014.

[109] G. Draisma, H. Dress, A. Ferreira, and L. De Haan, “Bivariate tail estimation: Dependence in asymptotic independence,” *Bernoulli*, pp. 251–280, 2004.

[110] Y. Goegebeur and A. Guillou, “Asymptotically unbiased estimation of the coefficient of tail dependence,” *Scandinavian Journal of Statistics*, vol. 40, no. 1, pp. 174–189, 2013.

[111] H. Manner and J. Segers, “Tails of correlation mixtures of elliptical copulas,” *Insurance: Mathematics and Economics*, vol. 48, no. 1, pp. 153–160, 2011.

[112] L. Li, K. Yuen, and J. Yang, “Distorted mix method for constructing copulas with tail dependence,” *Insurance: Mathematics and Economics*, vol. 57, pp. 77–89, 2014.

[113] W. Hoeffding, “A class of statistics with asymptotically normal distribution,” *The annals of mathematical statistics*, pp. 293–325, 1948.

[114] R. J. Serfling, *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009, vol. 162.